

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE
APPLIQUÉES

PAR
MARYSE CHAMPAGNE

OUTILS NON-PARAMÉTRIQUES DE DÉTECTION DE
RUPTURES UNIVARIÉES ET DE DÉPENDANCE :
APPLICATIONS EN HYDROLOGIE

SEPTEMBRE 2009

©Maryse Champagne, 2009

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

REMERCIEMENTS

Je tiens ici à remercier toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce mémoire. Tout d'abord, je tiens à remercier mon directeur de recherche, Jean-François Quessy, professeur au département de mathématiques et d'informatique de l'Université du Québec à Trois-Rivières. En plus de m'avoir proposé ce projet passionnant, il a su m'encourager et me soutenir tout au long de ce travail. La confiance qu'il m'a témoignée, ainsi que sa grande disponibilité, ont assuré la réussite de ce projet.

Je suis également reconnaissante envers ma co-directrice de recherche, Anne-Catherine Favre, professeure au Centre Eau, Terre et Environnement de l'Institut national de la recherche scientifique. J'ai grandement apprécié travailler avec elle, d'autant plus que ceci m'a permis de bénéficier de ses judicieux conseils. Un merci particulier va à Mériem Saïd, assistante de recherche, qui m'a encouragée et guidée tout au long du stage que j'ai effectué à l'été 2008 au Centre Eau, Terre et Environnement de l'Institut national de la recherche scientifique. Ce stage m'a été grandement bénéfique.

Je désire, bien sûr, remercier les membres de ma famille pour m'avoir toujours encouragé durant mes études. Je souhaite également souligner le support de mes amis, dont la présence m'a permis de me divertir, même dans les moments plus difficiles.

Je remercie également Alain Chalifour, Anne-Catherine Favre et Mériem Saïd d'avoir accepté de faire parti du jury d'évaluation de mon mémoire. Leurs judicieux commentaires ont permis d'améliorer la qualité de mon travail.

Une partie des travaux décrits dans ce mémoire a été financée par des octrois individuels accordés à M. Jean-François Quessy et Mme Anne-Catherine Favre par le *Conseil de Recherche en Sciences Naturelles et en Génie du Canada*. Je remercie également l'*Institut des Sciences Mathématiques* du Québec pour les bourses d'étude qu'il m'a accordées.

Table des matières

Remerciements	i
Liste des tableaux	vii
Liste des figures	xi
Chapitre 1. Introduction	1
1.1 Détection d’une rupture dans une loi unidimensionnelle	1
1.2 Détection d’une rupture dans une loi multidimensionnelle . . .	6
1.3 Structure du mémoire et objectifs	7
Chapitre 2. Monte-Carlo study of breakpoint detection tests in Lombard’s smooth-change model	10
2.1 Introduction	10
2.2 Preliminaries	12
2.3 Power study	17
2.3.1 Size of the tests	17
2.3.2 Breakpoint detection in the mean	18
2.3.3 Breakpoint detection in the variance	19
2.3.4 Breakpoint detection under simultaneous changes in the mean and variance	20
2.4 Robustness	20

2.4.1	Breakpoint detection in the mean under a smooth change in the variance	21
2.4.2	Breakpoint detection in the variance under a smooth change in the mean	21
2.5	Behavior under serial dependence	22
2.6	Discussion	24
Chapitre 3. Modélisation de phénomènes par les copules		35
3.1	Quelques considérations théoriques sur les copules	36
3.2	Mesure et détection de la dépendance	41
3.2.1	Mesures de dépendance	42
3.2.2	Détection graphique de la dépendance	44
3.3	Modélisation de données multidimensionnelles par les copules .	47
3.3.1	Estimation des paramètres d'une copule	48
3.3.2	Tests d'adéquation	51
Chapitre 4. Nouveaux tests de détection d'un point de rupture dans la dépendance		56
4.1	Problématique	56
4.2	Test de Dias & Embrechts (2004)	58
4.3	Test de Gombay & Horváth (1999)	60
4.4	Description d'une nouvelle méthodologie	64
4.4.1	Contexte	64
4.4.2	Stabilisation des marges : en théorie	66
4.4.3	Stabilisation des marges : en pratique	69
4.4.4	Identification des points de rupture marginaux	70
4.4.5	Étapes de conduite du test	71

Chapitre 5. Évaluation de la puissance et de la robustesse des nouveaux tests	74
5.1 Puissance de détection de points de rupture dans la dépendance sous des marges exemptes de ruptures	75
5.1.1 Rupture dans le type de dépendance des copules pour un même degré de dépendance	77
5.1.2 Rupture au niveau du degré de dépendance des copules pour un même type de dépendance	79
5.1.3 Rupture simultanée dans le type de dépendance ainsi qu'au niveau du degré de dépendance des copules . . .	80
5.2 Puissance de détection de points de rupture dans la dépendance sous des ruptures dans les marges	82
5.2.1 Description des scénarios considérés	82
5.2.2 Résultats et interprétations	85
5.3 Conclusion générale sur l'étude de puissance	89
Chapitre 6. Application sur des données hydrologiques réelles	90
6.1 Présentation des données	90
6.2 Détection de ruptures pour les bassins <i>À la Baleine</i> et <i>Georges</i>	92
6.2.1 Analyses descriptives des échantillons	93
6.2.2 Détection de points de rupture dans les marges	95
6.2.3 Détection de ruptures dans la forme de la dépendance .	99
6.3 Analyse des ruptures dans la dépendance pour quelques bassins versants du Québec	101
6.3.1 Détection des ruptures dans les marges	102
6.3.2 Analyse des points de rupture dans la dépendance . . .	104

6.4 Sommaire des analyses	106
Conclusions et perspectives	108
Annexe A. Programmes en Matlab	110
Références	118

LISTE DES TABLEAUX

2.1	Estimated critical values of the tests based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} at the 5% level (10 000 replicates)	15
2.2	Size of the tests based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} , i.e. percentage of rejection of the null hypothesis whenever \mathcal{H}_0 is true, estimated with the help of 10 000 replicates from the Normal distribution with unit variance	18
2.3	Power, estimated with the help of 10 000 replicates, of S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} to detect departures from \mathcal{H}_0 for i.i.d. sequences from the Normal distribution with unit variance, $n = 50$	26
2.4	Power, estimated with the help of 10 000 replicates, of S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} to detect departures from \mathcal{H}_0 for i.i.d. sequences from the Normal distribution with unit variance, $n = 100$	27
2.5	Power, estimated with the help of 10 000 replicates, of S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} to detect breakpoints in the variance for i.i.d. sequences of size $n = 50$ from the Normal distribution with zero mean	28
2.6	Power, estimated with the help of 10 000 replicates, of S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} to detect breakpoints in the variance for i.i.d. sequences of size $n = 100$ from the Normal distribution with zero mean	29

2.7	Power, estimated with the help of 10 000 replicates, of S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} to detect departures from \mathcal{H}_0 for i.i.d. sequences from the Exponential distribution. Initial mean : $\lambda_1 = 1$; sample size : $n = 50$	30
2.8	Size of the tests of breakpoint detection in the mean based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} under i.i.d. sequences from the Normal distribution with mean 0 and a smooth change in the variance.	31
2.9	Size of the tests of breakpoint detection in the variance based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} under i.i.d. sequences from the Normal distribution with variance 1 and a smooth change in the mean, $n = 50$	32
2.10	Size of the tests of breakpoint detection in the mean based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} under serial dependence, $n = 100$	33
2.11	Power of the tests of breakpoint detection in the mean based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} under serial dependence based on Clayton and $ \mu_1 - \mu_2 = 0.25$, $n = 100$	34
5.1	Description de trois copules Archimédiennes	76
5.2	Puissance, estimée par 1000 répétitions, du test de détection de rupture de copules pour des changements dans le type de dépendance mais pour un tau de Kendall stable. Panneau supérieur : $n = 50$; panneau inférieur : $n = 100$	78

5.3	Puissance, estimée par 1000 répétitions, du test de détection de rupture de copules pour des changements dans le tau de Kendall mais sous des modèles de copules stables. Panneau supérieur : $n = 50$; panneau inférieur : $n = 100$	80
5.4	Puissance, estimée par 1000 répétitions, du test de détection de rupture de copules pour des changements dans le tau de Kendall et la copule. Panneau supérieur : $n = 50$; panneau inférieur : $n = 100$	82
5.5	Puissance de la procédure complète pour la détection de points de rupture de copules sous le scénario de référence ainsi que sous les scénarios A et B pour des tailles d'échantillons de $n = 50$	87
5.6	Puissance de la procédure complète pour la détection de points de rupture de copules sous le scénario C pour des tailles d'échantillons de $n = 50$	88
6.1	Statistiques descriptives des écoulements annuels moyens sur la rivière <i>À la Baleine</i> entre les années 1958 et 2099	93
6.2	Statistiques descriptives des écoulements annuels moyens sur la rivière <i>George</i> entre les années 1958 et 2099	95
6.3	Tests de détection de rupture en moyenne et en variance pour les bassins des rivières <i>À la Baleine</i> et <i>George</i>	101
6.4	Statistiques S_{n_1} et années de rupture en moyenne ($\hat{\mathcal{K}}_{\hat{\mu}_1}, \hat{\mathcal{K}}_{\hat{\mu}_2}$) et en variance ($\hat{\mathcal{K}}_{\hat{\sigma}_1}, \hat{\mathcal{K}}_{\hat{\sigma}_2}$) des vingt et un bassins versants du Québec	103

- 6.5 Statistiques de test et valeurs critiques estimées (triangle supérieur)
et années de rupture de la dépendance, s'il y a lieu (triangle
inférieur) pour les bassins versants du centre-ouest du Québec 104
- 6.6 Statistiques de test et valeurs critiques estimées (triangle supérieur)
et années de rupture de la dépendance, s'il y a lieu (triangle
inférieur) pour les bassins versants du centre-est du Québec . . 106

LISTE DES FIGURES

1.1	Illustration d'une rupture abrupte de moyenne	2
1.2	Illustration d'une rupture graduelle de moyenne	3
6.1	Bassins versants du Québec	92
6.2	Écoulements annuels moyens sur la rivière <i>À la Baleine</i> entre les années 1958 et 2099	94
6.3	Droite de Henry des écoulements annuels moyens sur la rivière <i>À la Baleine</i> entre les années 1958 et 2099	95
6.4	Écoulements annuels moyens sur la rivière <i>George</i> entre les années 1958 et 2099	96
6.5	Droite de Henry des écoulements annuels moyens sur la rivière <i>George</i> entre les années 1958 et 2099	97
6.6	Nuage de points des rangs normalisés des écoulements annuels moyens des rivières <i>À la Baleine</i> et <i>George</i>	98
6.7	Écoulements annuels moyens sur la rivière <i>À la Baleine</i> entre les années 1958 et 2099 après stabilisation de la moyenne . . .	99
6.8	Écoulements annuels moyens sur la rivière <i>George</i> entre les années 1958 et 2099 après stabilisation de la moyenne	100
6.9	Outil graphique pour détecter une rupture de copule, illustra- tion pour les bassins BAL et GEO	102
6.10	Détection graphique de la dépendance pour le réservoir <i>Ca-</i> <i>niapiscou</i> et le réservoir <i>Manic5</i>	105
6.11	Détection graphique de la dépendance pour les rivières <i>Moisie</i> (MOI) et <i>Romaine</i> (ROM)	106

CHAPITRE 1

INTRODUCTION

Une rupture dans une série de variables aléatoires se définit comme un changement, au cours du temps, de leur loi de probabilité. Il existe plusieurs types de ruptures moins forts que les modifications dans la loi. On pense, entre autres, aux changements dans la moyenne ou dans la variance. On peut classer les procédures de détection de ruptures en deux catégories : (i) les tests pour les lois unidimensionnelles et (ii) les tests pour les lois multidimensionnelles. Ces deux catégories sont décrites aux sections 1.1 et 1.2, alors que la section 1.3 donne un aperçu de la structure du mémoire.

1.1 Détection d'une rupture dans une loi unidimensionnelle

Soit une suite de variables aléatoires indépendantes X_1, \dots, X_n . On suppose que les fonctions de répartition respectives sont $F_1(x), \dots, F_n(x)$, où $F_i(x) = P(X_i \leq x)$. On dit que cette suite possède une *rupture de loi* au temps \mathcal{K} si $F_1(x) = \dots = F_{\mathcal{K}}(x) \neq F_{\mathcal{K}+1}(x) = \dots = F_n(x)$.

Quand les observations X_1, \dots, X_n sont supposées provenir d'un modèle paramétrique $F(x; \theta)$, alors les ruptures potentielles peuvent survenir uniquement via un changement du paramètre θ . Autrement dit, on a $F_i(x) = F(x; \theta_i)$, $1 \leq i \leq n$. Ainsi, on parle d'une *rupture de paramètre* au temps \mathcal{K} si $\theta_1 = \dots = \theta_{\mathcal{K}} \neq \theta_{\mathcal{K}+1} = \dots = \theta_n$. Le paramètre θ peut représenter, par exemple, la moyenne ou la variance d'une loi de probabilité.

Pour illustrer ce qui vient d'être mentionné, prenons une série de $n = 100$ observations tirées de la loi normale. On suppose une rupture abrupte de moyenne au temps $\mathcal{K} = 25$. Plus précisément, on a

$$X_1, \dots, X_{25} \sim \mathcal{N}(0, 1) \quad \text{et} \quad X_{26}, \dots, X_{100} \sim \mathcal{N}(3, 1),$$

où $X \sim \mathcal{N}(\mu, \sigma^2)$ signifie que la loi de la variable aléatoire X est la normale de moyenne μ et de variance σ^2 . La figure 1.1 présente une réalisation de ce modèle. On peut apercevoir un changement de moyenne aux environs de la 25-ème observation.

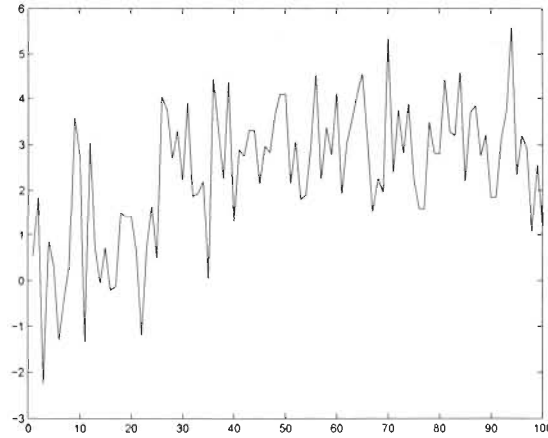


Figure 1.1: Illustration d'une rupture abrupte de moyenne

Plutôt que d'être abrupte, comme à la figure 1.1, une rupture peut aussi se produire de façon progressive. Pour illustrer, supposons $n = 100$ observations tirées d'une distribution normale dont la moyenne subit une modification graduelle entre les observations $\mathcal{K}_1 = 25$ et $\mathcal{K}_2 = 60$. Spécifiquement, on a

$$X_1, \dots, X_{25} \sim \mathcal{N}(0, 1) \quad \text{et} \quad X_{60}, \dots, X_{100} \sim \mathcal{N}(3, 1),$$

alors que la moyenne entre les temps $\mathcal{K}_1 = 25$ et $\mathcal{K}_2 = 60$ augmente de façon linéaire pour passer de 0 à 3. La figure 1.2 présente les données simulées selon ce modèle simple.

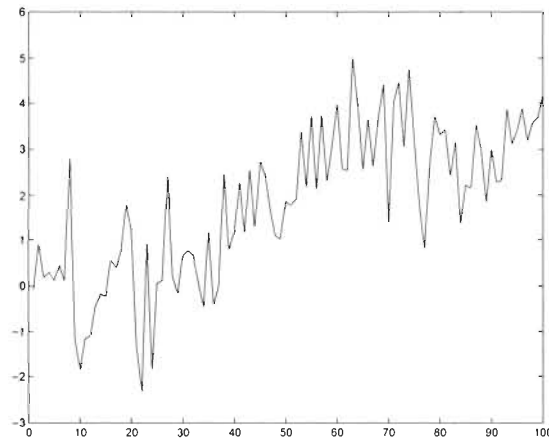


Figure 1.2: Illustration d'une rupture graduelle de moyenne

Le graphique semble assez explicite quant à la présence d'une rupture dans la moyenne. En pratique, les conclusions ne sont pas toujours aussi faciles à tirer. Il est donc primordial de disposer de tests statistiques afin de détecter formellement des ruptures, qu'elles soient abruptes ou progressives.

Dans ce qui suit, une brève revue de littérature sur les tests de détection de ruptures dans les lois unidimensionnelles est présentée. Ces tests peu-

vent se classer en deux catégories, à savoir (i) les approches paramétriques et (ii) les approches non-paramétriques. Pour les méthodes de type (i), une hypothèse est émise quant à la forme de la distribution des observations. Cette supposition permet généralement le calcul explicite de la distribution d'échantillonnage de la statistique de test. Cependant, une mauvaise identification de la loi sous-jacente peut mener à de graves erreurs d'interprétation. Les méthodes de type (ii) ne souffrent pas de cette contrainte, car aucune hypothèse n'est émise concernant la forme de la loi. En l'absence d'information fiable sur la forme de la distribution, ces approches sont à privilégier.

Les approches paramétriques ont fait l'objet de nombreux travaux, et ce depuis fort longtemps. Le contexte le plus simple consiste à supposer que les observations proviennent de la loi normale. Plusieurs méthodologies ont été développées sous ce paradigme. Se référer à Chen & Gupta (1999) pour une revue des tests de détection et des méthodes d'estimation de points de rupture pour des modèles basés sur la distribution normale. Une autre approche populaire consiste à utiliser le test du rapport de vraisemblances. De cette façon, il est possible de faire les tests en supposant que les observations proviennent d'une loi autre que la normale. À ce sujet, voir les articles de Hinkley (1970) et Hinkley (1972). Une contribution récente provient de Zou et al. (2007). Aussi, plusieurs approches sont de type *somme cumulée* (CUSUM). On peut citer la contribution de Horváth & Steinebach (2000), où les ruptures en moyenne et en variance sont considérées. Antoch & Hušková (2001), quant à eux, ont proposé des tests de permutation pour les paramètres de localisation et d'échelle d'une loi. Un test basé sur les fonctions caractéristiques empiriques a été développé par Hušková & Meintanis (2006).

Plusieurs tests ont également été conçus afin de tenir compte de la possible dépendance sérielle. Par exemple, Gombay (2008) a proposé des tests de détection de ruptures dans un modèle auto-régressif d'ordre p . Ces procédures permettent de tester pour des changements abrupts de la moyenne, de la variance et des paramètres auto-régressifs. Berkes et al. (2009) ont utilisé une procédure de type CUSUM pondéré afin d'identifier une rupture éventuelle au niveau de la moyenne et de la structure de covariance de processus linéaires. Des applications de quelques tests de détection de ruptures sur des données météorologiques et hydrologiques sont offertes par Jarušková (1997).

La littérature concernant la détection de ruptures par des approches non paramétriques semble remonter aussi loin qu'à Page (1954). Cet auteur a employé une version de la méthode CUSUM. Pettitt (1979) a proposé un test pour identifier des changements abrupts de moyenne. Ce test a été généralisé par l'apport important de Lombard (1987), qui a considéré un modèle beaucoup plus général (et réaliste) permettant de possibles changements graduels. Les statistiques de test qu'il a proposées servent autant à la détection de ruptures de moyenne que de variance. Une revue de littérature sur plusieurs procédures non-paramétriques, ainsi que des études de puissance, sont présentées par Scariano & Watkins (1988) et Bryden et al. (1995).

1.2 Détection d'une rupture dans une loi multidimensionnelle

L'étude des points de rupture dans une série d'observations multivariées a été abordée par quelques auteurs. Par souci de simplicité, nous considérons le cas à deux variables. La généralisation à $d \geq 2$ variables est directe. Soit donc une suite de vecteurs aléatoires indépendants $(X_1, Y_1), \dots, (X_n, Y_n)$ tels que $P(X_i \leq x, Y_i \leq y) = H_i(x, y)$, où H_1, \dots, H_n sont des fonctions de répartition bivariées. Dire qu'il s'est produit une rupture au temps \mathcal{K} revient à affirmer que $H_1 = \dots = H_{\mathcal{K}} \neq H_{\mathcal{K}+1} = \dots = H_n$.

La littérature concernant la détection de points de rupture dans les lois multidimensionnelles est moins étoffée que pour le cas unidimensionnelle. Néanmoins, on peut citer les contributions de Csörgö & Horváth (1988), Gombay & Horváth (1999) et Horváth et al. (1999). Un problème majeur dans l'utilisation de ces approches tient à l'impossibilité d'identifier la source de la rupture. Ainsi, comme $F_i(x) = P(X_i \leq x) = H_i(x, +\infty)$ et $G_i(y) = P(Y_i \leq y) = H_i(+\infty, y)$, un changement survenu uniquement dans la distribution de la suite X_1, \dots, X_n (ou de façon similaire, uniquement dans celle de la suite Y_1, \dots, Y_n) induit un changement observé dans les lois conjointes H_1, \dots, H_n . Il serait intéressant d'avoir une approche qui permette de détecter, indépendamment, des changements pouvant se produire dans les séries univariées X_1, \dots, X_n et Y_1, \dots, Y_n , des changements pouvant apparaître dans le lien qui unit les variables aléatoires X_i et Y_i , $i \in \{1, \dots, n\}$.

Une solution à ce problème réside dans l'utilisation des copules. La principale

utilité de cette approche est qu'elle permet de décomposer le comportement d'un couple de variables aléatoires (X, Y) en trois sources distinctes : (i) le comportement de X , (ii) le comportement de Y et (iii) la dépendance entre X et Y . Ainsi, en utilisant les copules, il sera possible de détecter des ruptures à la fois dans les lois unidimensionnelles et dans la structure de dépendance.

La seule contribution connue sur la détection d'une rupture de copule provient de Dias & Embrechts (2004). Toutefois, leur approche suppose que les comportements marginaux des variables sont connus et que le modèle de copules est spécifié, à un paramètre près. Ces hypothèses sont beaucoup trop fortes pour les applications. Il faut mentionner que la méthode développée par Gombay & Horváth (1999) permet de tester pour un changement dans la copule, en autant qu'une adaptation judicieuse en est faite. Ce sera un des sujets principaux de ce mémoire.

1.3 Structure du mémoire et objectifs

Ce mémoire poursuit deux objectifs :

1. Étudier la puissance des tests de détection de rupture de moyenne et de variance de Lombard (1987) dans une série unidimensionnelle ;
2. Proposer une méthodologie pour la détection de rupture d'un modèle de copule dans une série multidimensionnelle.

Le premier objectif est poursuivi au chapitre 2. Cette partie est entièrement constituée d'un article scientifique, rédigé en anglais, dont le titre est *Monte-Carlo study of breakpoint detection tests in Lombard's smooth-change model*. L'article présente une étude exhaustive de puissance et de robustesse pour six procédures de test proposées par Lombard (1987) pour la détection de ruptures dans une série unidimensionnelle. De nombreux scénarios pouvant se produire dans la nature sont étudiés, ce qui permet d'identifier les meilleures statistiques à utiliser dans un contexte donné. Ces résultats, d'un intérêt indépendant, seront également utiles dans les chapitres subséquents concernant les ruptures de lois multidimensionnelles.

Les chapitres suivants concernent l'atteinte du deuxième objectif. Il s'agit de développer un nouveau test de détection d'un changement dans la structure de dépendance d'une série de vecteurs aléatoires. La dépendance est considérée à travers la théorie des copules. Cela permet de séparer les liens d'interdépendance des effets individuels dans un vecteur aléatoire. La procédure proposée est entièrement non paramétrique au sens où aucun modèle de copules n'est préalablement supposé. Les hypothèses sur les comportements marginaux sont également relâchées. Cette procédure permettra de tester formellement la présence ou non de points de rupture dans la structure de dépendance. Il s'agira de modifier judicieusement le test de Gombay & Horváth (1999) afin de le rendre applicable à notre contexte.

Spécifiquement, le chapitre 3 présente une introduction à la théorie des copules. Des exemples, ainsi que quelques mesures de dépendance et des méthodes d'estimation, y sont présentés. Le chapitre 4 s'intéresse aux tests

de détection de ruptures dans la dépendance. On y retrouve la description des tests de Dias & Embrechts (2004) et de Gombay & Horváth (1999) qui considèrent des cas où les comportements marginaux sont stables. Par la suite, l'apport innovateur principal de ce travail de recherche est introduit, c'est-à-dire la méthodologie de notre nouveau test. Cette méthode permet de détecter des ruptures dans la dépendance tout en permettant aux marges d'être sujettes à des ruptures. Au chapitre 5, nous présentons les résultats d'une étude de simulation effectuée afin d'attester de la validité et de la puissance du nouveau test. Le chapitre 6 contient des applications concrètes sur des données simulées issues du *Modèle Régional Canadien du Climat*.

CHAPITRE 2

MONTE-CARLO STUDY OF BREAKPOINT DETECTION TESTS IN LOMBARD'S SMOOTH-CHANGE MODEL

Mots-clés: Détection de points de rupture, étude de puissance, robustesse, modèle à changements graduels

Résumé: La puissance de six statistiques de test pour la détection de points de rupture dans la moyenne et la variance d'une chronique est étudiée. Ces statistiques furent proposées par Lombard (1987) pour détecter des écarts à l'hypothèse nulle de stabilité dans un modèle à changements graduels. La robustesse des procédures en présence d'hétérogénéité et de dépendance sérielle est également considérée.

2.1 Introduction

Many phenomena that one observes are subject to structural changes over time. More specifically, if X_1, \dots, X_n are observations taken at equal time intervals, it may happen that the distribution of $X_1, \dots, X_{\mathcal{K}}$ differs from that of $X_{\mathcal{K}+1}, \dots, X_n$. In other words, an abrupt change at time \mathcal{K} can occur.

An extensive literature exists on the problem of change-point detection. Many of the contributions are parametric in the sense that a specific model is supposed. In that context, the alternative hypothesis is stated in terms of a changing in a given parameter. See, for instance, Jandhyala et al. (2000), Perreault et al. (2000a) and Perreault et al. (2000b). See also Jandhyala et al. (2002) and Fotopoulos & Jandhyala (2007).

Since a bad specification of the parametric model can lead to misleading conclusions, nonparametric tests are often advisable. This is the approach taken by Carlstein (1988) and Ferger (1994, 1995), where nonparametric tests for breakpoints in distribution functions are proposed. For abrupt changes in moments, nonparametric tests exist, e.g. Hsu (1977), Pettitt (1979) and Abd-Rabou & Gad (2007). If $\theta_1, \dots, \theta_n$ are the values of a given parameter of interest for independent observations X_1, \dots, X_n , these tests assume an abrupt-change model of the form

$$\theta_i = \begin{cases} \xi_1, & 1 \leq i \leq \mathcal{K}; \\ \xi_2, & \mathcal{K} < i \leq n, \end{cases} \quad (2.1)$$

where \mathcal{K} is the breakpoint. However, this assumption of an abrupt change is often unrealistic in many hydrological and financial situations. To this end, Lombard (1987) proposed the smooth-change model

$$\theta_i = \begin{cases} \xi_1, & 1 \leq i \leq \mathcal{K}_1; \\ \xi_1 + \frac{(i - \mathcal{K}_1)(\xi_2 - \xi_1)}{\mathcal{K}_2 - \mathcal{K}_1}, & \mathcal{K}_1 < i \leq \mathcal{K}_2 \\ \xi_2, & \mathcal{K}_2 < i \leq n. \end{cases} \quad (2.2)$$

One obviously recovers model (2.1) in the special case $\mathcal{K}_1 = \mathcal{K}_2 = \mathcal{K}$. Note that other models that generalize the abrupt-change model have been pro-

posed, e.g. the epidemic alternative model. See Antoch & Hušková (1996), for instance.

In this paper, the efficiency and robustness of test statistics to detect structural changes of the type (2.2) will be investigated. Both the cases of breakpoint detection in the mean and in the variance will be considered. The test statistics that will be investigated have been proposed by Lombard (1987). This study complements that of Bryden et al. (1995), who considered only the abrupt-change model in their simulations. Furthermore, the robustness of the tests when some basic assumptions are violated is considered.

The article is organized as follows : in Section 2, the six test statistics that will be investigated are presented. Section 3 investigates the power of the tests under breakpoints in the mean and in the variance. Sections 4–5 study the robustness of the tests under two general scenarios : (i) heterogeneity of a nuisance parameter, for example when the variance changes when one is testing for a change in the mean and (ii) serial dependence. A discussion section ends the paper.

2.2 Preliminaries

Let R_i be the normalized rank of X_i among X_1, \dots, X_n , i.e. $R_i = F_n(X_i)$, where

$$F_n(x) = \frac{1}{n+1} \sum_{\ell=1}^n \mathbb{I}(X_\ell \leq x) \quad (2.3)$$

is the re-scaled empirical distribution function. The function \mathbb{I} is the usual indicator function of a set. Consider a score function ϕ such that

$$0 < \int_0^1 \phi^2(u) \, du < \infty.$$

Define the rank score of X_i by

$$Z_i = \frac{\phi(R_i) - \bar{\phi}}{\sigma_\phi},$$

where

$$\bar{\phi} = \frac{1}{n} \sum_{i=1}^n \phi\left(\frac{i}{n+1}\right) \quad \text{and} \quad \sigma_\phi^2 = \frac{1}{n} \sum_{i=1}^n \left\{ \phi\left(\frac{i}{n+1}\right) - \bar{\phi} \right\}^2.$$

Let

$$W_j = \sum_{\ell=1}^j Z_\ell$$

and

$$\mathcal{L}_{\mathcal{K}_1, \mathcal{K}_2} = \sum_{j=\mathcal{K}_1+1}^{\mathcal{K}_2} W_j. \quad (2.4)$$

Lombard (1987) proposed three Cramér–von Mises type test statistics based on $\mathcal{L}_{\mathcal{K}_1, \mathcal{K}_2}$, namely

$$S_{n1} = \frac{1}{n^5} \sum_{1 \leq \mathcal{K}_1 < \mathcal{K}_2 \leq n} \mathcal{L}_{\mathcal{K}_1, \mathcal{K}_2}^2, \quad (2.5)$$

$$S_{n2} = \frac{1}{n^4} \sum_{\mathcal{K}_1=1}^n \mathcal{L}_{\mathcal{K}_1, n}^2, \quad (2.6)$$

$$S_{n3} = \frac{1}{n^2} \sum_{\mathcal{K}_1=1}^{n-1} \mathcal{L}_{\mathcal{K}_1, \mathcal{K}_1+1}^2. \quad (2.7)$$

These statistics are suitable respectively for the three following models :

(\mathcal{M}_1) the general *smooth-change* model, i.e. $1 < \mathcal{K}_1 < \mathcal{K}_2 < n$;

(\mathcal{M}_2) the *trend-change* model, i.e. $\mathcal{K}_2 = n$;

(\mathcal{M}_3) the *abrupt-change* model, i.e. $\mathcal{K}_2 = \mathcal{K}_1 + 1$.

Kolmogorov–Smirnov versions of S_{n1} , S_{n2} and S_{n3} are

$$T_{n1} = n^{-3/2} \max_{1 \leq \mathcal{K}_1 < \mathcal{K}_2 \leq n} |\mathcal{L}_{\mathcal{K}_1, \mathcal{K}_2}|, \quad (2.8)$$

$$T_{n2} = n^{-3/2} \max_{1 \leq \mathcal{K} \leq n} |\mathcal{L}_{\mathcal{K}, n}|, \quad (2.9)$$

$$T_{n3} = n^{-1/2} \max_{1 \leq \mathcal{K} < n} |\mathcal{L}_{\mathcal{K}, \mathcal{K}+1}|. \quad (2.10)$$

Lombard (1987) obtained that

$$\begin{aligned} S_{n1} \rightsquigarrow S_1 &= \int_0^1 \int_u^1 \left\{ \int_u^v \beta(t) dt \right\}^2 dv du, \\ S_{n2} \rightsquigarrow S_2 &= \int_0^1 \left\{ \int_u^1 \beta(t) dt \right\}^2 du, \\ S_{n3} \rightsquigarrow S_3 &= \int_0^1 \{\beta(t)\}^2 dt, \end{aligned}$$

where \rightsquigarrow means *convergence in distribution*. Here, $\{\beta(t), 0 \leq t \leq 1\}$ is a Brownian bridge, i.e. a continuous centered Gaussian process on $[0, 1]$ with $E\{\beta(s)\beta(t)\} = \min(s, t) - st$. Lombard (1987) then used Karh nen–Lo  ve expansions to deduce the asymptotic critical values of the tests based on S_{n1} , S_{n2} and S_{n3} . While expressions for the limiting distributions of T_{n1} , T_{n2} and T_{n3} can be obtained in terms of functionals of β , the significance points of the tests based on the latter will have to be estimated via Monte-Carlo simulations. Some values are reported in Table 2.1.

In this paper, the Wilcoxon score function, i.e. $\phi_1(u) = 2u - 1$, will be used to detect changes in the mean, while the Mood score $\phi_2(u) = (2u - 1)^2$ will be used for changes in the variance.

Table 2.1: Estimated critical values of the tests based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} at the 5% level (10 000 replicates)

n	S_{n1}	S_{n2}	S_{n3}	T_{n1}	T_{n2}	T_{n3}
50	0.0417	0.1212	0.4659	0.5752	0.5726	1.2757
100	0.0398	0.1201	0.4530	0.5672	0.5653	1.2995
150	0.0402	0.1222	0.4533	0.5651	0.5626	1.3080
∞	0.0403	0.1242	0.4614	0.5616*	0.5588*	1.3277*

* : Estimated with $n = 1000$

Proposition 2.1 justifies the choice of breakpoints that are considered in the upcoming numerical investigations.

Proposition 2.1. *Let X_1, \dots, X_n follow the smooth-change model (2.2). Then the test statistics S_{nk} and T_{nk} , $k = 1, 2, 3$, are invariant under the following situations :*

(i) Y_1, \dots, Y_n , where $Y_i = -X_i$, $1 \leq i \leq n$;

(ii) Y_1, \dots, Y_n , where $Y_i = X_{n+1-i}$, $1 \leq i \leq n$.

Proof. Let $R_i = F_n(X_i)$ and $\tilde{R}_i = \tilde{F}_n(Y_i)$, where F_n is described in (2.3) and

$$\tilde{F}_n(y) = \frac{1}{n+1} \sum_{\ell=1}^n \mathbb{I}(Y_\ell \leq y).$$

(i) First of all, note that

$$\begin{aligned}
\tilde{R}_i = \tilde{F}_n(Y_i) &= \frac{1}{n+1} \sum_{\ell=1}^n \mathbb{I}(-X_\ell \leq -X_i) \\
&= \frac{1}{n+1} \sum_{\ell=1}^n \{1 - \mathbb{I}(X_\ell < X_i)\} \\
&= \frac{n}{n+1} - \frac{1}{n+1} \sum_{\ell=1}^n \mathbb{I}(X_\ell < X_i) \\
&= \frac{n}{n+1} - \frac{1}{n+1} \sum_{\ell=1}^n \mathbb{I}(X_\ell \leq X_i) + \frac{1}{n+1} = 1 - R_i.
\end{aligned}$$

Hence, since $\bar{\phi}_1 = 0$ and $\phi_1(\tilde{R}_i) = \phi_1(1 - R_i) = -\phi_1(R_i)$, one has $\tilde{Z}_i = -Z_i$.

This yields

$$\tilde{W}_j = \sum_{\ell=1}^j \tilde{Z}_\ell = -W_j \quad \text{and} \quad \tilde{\mathcal{L}}_{\mathcal{K}_1, \mathcal{K}_2} = \sum_{j=\mathcal{K}_1+1}^{\mathcal{K}_2} \tilde{W}_{n,j} = -\mathcal{L}_{\mathcal{K}_1, \mathcal{K}_2},$$

and one can conclude that $|\tilde{\mathcal{L}}_{\mathcal{K}_1, \mathcal{K}_2}| = |\mathcal{L}_{\mathcal{K}_1, \mathcal{K}_2}|$. The same conclusion holds for ϕ_2 since $\phi_2(\tilde{R}_i) = \phi_2(1 - R_i) = \phi_2(R_i)$, and thus $\tilde{Z}_i = Z_i$.

(ii) One has

$$\begin{aligned}
\tilde{R}_i = \tilde{F}_n(Y_i) &= \frac{1}{n+1} \sum_{\ell=1}^n \mathbb{I}(X_{n+1-\ell} \leq X_{n+1-i}) \\
&= \frac{1}{n+1} \sum_{\ell=1}^n \mathbb{I}(X_\ell \leq X_{n+1-i}) = R_{n+1-i},
\end{aligned}$$

so that $\tilde{Z}_i = Z_{n+1-i}$. Hence,

$$\tilde{W}_j = \sum_{\ell=1}^j Z_{n+1-\ell} = \sum_{\ell=n+1-j}^n Z_\ell = \sum_{\ell=1}^n Z_\ell - \sum_{\ell=1}^{n-j} Z_\ell = -\sum_{\ell=1}^{n-j} Z_\ell = -W_{n-j},$$

where the fact that

$$\sum_{\ell=1}^n Z_\ell = \sum_{\ell=1}^n \left\{ \frac{\phi(R_\ell) - \bar{\phi}}{\sigma_\phi} \right\} = \frac{1}{\sigma_\phi} \sum_{\ell=1}^n \left\{ \phi\left(\frac{\ell}{n+1}\right) - \bar{\phi} \right\} = 0$$

has been used. Thus

$$\tilde{\mathcal{L}}_{\mathcal{K}_1, \mathcal{K}_2} = \sum_{j=\mathcal{K}_1+1}^{\mathcal{K}_2} \tilde{W}_j = - \sum_{j=\mathcal{K}_1+1}^{\mathcal{K}_2} W_{n-j} = - \sum_{j=n-\mathcal{K}_2}^{n+1-\mathcal{K}_1} W_j = -\mathcal{L}_{n-\mathcal{K}_2, n+1-\mathcal{K}_1}.$$

Finally,

$$\begin{aligned} \left\{ \tilde{\mathcal{L}}_{\mathcal{K}_1, \mathcal{K}_2}; 1 \leq \mathcal{K}_1 < \mathcal{K}_2 \leq n \right\} &= \left\{ -\mathcal{L}_{n-\mathcal{K}_2, n+1-\mathcal{K}_1}; 1 \leq \mathcal{K}_1 < \mathcal{K}_2 \leq n \right\} \\ &= \left\{ -\mathcal{L}_{\mathcal{K}_1, \mathcal{K}_2}; 1 \leq \mathcal{K}_1 < \mathcal{K}_2 \leq n \right\} \end{aligned}$$

entails that $\tilde{S}_{nk} = S_{nk}$ and $\tilde{T}_{nk} = T_{nk}$ for all $k = 1, 2, 3$.

2.3 Power study

2.3.1 Size of the tests

Before starting the investigation of the power of the tests, it is important to assess that the six procedures keep their nominal level under \mathcal{H}_0 . In Table 2.2, the probability of rejecting the null hypothesis under \mathcal{H}_0 is evaluated with the help of 10 000 random samples of size $n = 50, 100$ from the $\mathbb{N}(0, 1)$ distribution. The pre-assigned level of the tests is 5% and the critical values of Table 2.1 are used. One can see that all test statistics keep their nominal level very well, except for the Mood score $\phi_2(u) = (2u - 1)^2$ when $n = 50$, where a small bias is observed.

Table 2.2: Size of the tests based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} , i.e. percentage of rejection of the null hypothesis whenever \mathcal{H}_0 is true, estimated with the help of 10 000 replicates from the Normal distribution with unit variance

Score function	Sample size	S_{n1}	S_{n2}	S_{n3}	T_{n1}	T_{n2}	T_{n3}
Wilcoxon	50	4.55	4.68	4.77	4.79	4.79	5.07
ϕ_1	100	4.97	4.91	5.08	4.91	4.89	4.97
Mood	50	4.17	4.39	4.24	4.28	4.29	4.41
ϕ_2	100	5.09	4.90	5.10	5.10	5.11	4.50

2.3.2 Breakpoint detection in the mean

In Table 2.3 and Table 2.4, the power of the six test statistics for detecting a breakpoint in the mean is evaluated with $n = 50$ and $n = 100$, respectively. In that case, the Wilcoxon score function $\phi_1(u) = 2u - 1$ is employed. For that purpose, random samples from the Normal distribution with unit variance were generated. Note that the value of the latter is unimportant since the ranks are invariant under scale transformations. Six special cases of the smooth-change model (2.2) were considered, namely

- (i) $\mathcal{K}_1 = 25\%, \mathcal{K}_2 = 25\% + 1;$
- (ii) $\mathcal{K}_1 = 25\%, \mathcal{K}_2 = 50\%;$
- (iii) $\mathcal{K}_1 = 25\%, \mathcal{K}_2 = 75\%;$
- (iv) $\mathcal{K}_1 = 25\%, \mathcal{K}_2 = 100\%;$
- (v) $\mathcal{K}_1 = 50\%, \mathcal{K}_2 = 50\% + 1;$
- (vi) $\mathcal{K}_1 = 50\%, \mathcal{K}_2 = 100\%.$

[TABLE 2.3 AND 2.4 HERE]

2.3.3 Breakpoint detection in the variance

In Table 2.5 and Table 2.6, the power of the six test statistics with the Mood score $\phi_2(u) = (2u - 1)^2$ is evaluated with $n = 50$ and $n = 100$, respectively. The efficiency of the test to detect changes in the variance is investigated using random samples from the Normal distribution with a fixed mean, the latter being arbitrary since the ranks are invariant under location transformations. The same patterns as in Table 2.5–2.6 were considered for the location of the breakpoints.

[TABLE 2.5 AND 2.6 HERE]

2.3.4 Breakpoint detection under simultaneous changes in the mean and variance

It may occur, after the occurrence of a given event, that both the mean and the variance of a series change according to a pattern similar to the smooth-change model (2.2). This possibility can be modelled by the exponential law, whose mean and variance are linked through the same parameter, i.e. $E(X) = \text{var}(X) = \lambda$. Table 2.7 reports the estimations of power for the six test statistics using the Wilcoxon and Mood scores under random samples of size $n = 50$ from the exponential distribution. Each series were generated with the initial value $\lambda_1 = 1$, while λ_2 changes according to model (2.2) at the points $(\mathcal{K}_1, \mathcal{K}_2) = (25\%, 25\% + 1), (25\%, 50\%), (25\%, 100\%)$. The values of λ_2 are 1.5 and 3 respectively for the upper and lower panels.

[TABLE 2.7 HERE]

2.4 Robustness

When one tests for a change in some given parameter, it is assumed implicitly that the distributions of the observations only change in that particular parameter. This can be criticized in many situations where other characteristics of the series can also be affected.

2.4.1 Breakpoint detection in the mean under a smooth change in the variance

In Table 2.8, the ability of the tests of breakpoint detection in the mean to keep their nominal level under heterogeneity of the variance is studied. Random samples of size $n = 50, 100$ were generated from the $\mathbb{N}(0, \sigma_k^2)$, $k = 1, 2$, where σ_1^2 is the initial variance and σ_2^2 is the variance after a change has occurred according to model (2.2). The location of the changes are $(\mathcal{K}_1, \mathcal{K}_2) = (25\%, 25\% + 1), (25\%, 50\%), (25\%, 100\%)$.

[TABLE 2.8 HERE]

2.4.2 Breakpoint detection in the variance under a smooth change in the mean

An investigation similar to that in subsection 2.4.1 has been conducted for the tests of breakpoint detection in the variance under smooth changes in the mean. The results are reported in Table 2.9. The ability of the tests to keep their nominal level under heterogeneity of the mean is evaluated using random samples of size 100 from the $\mathbb{N}(\mu_k, 1)$ distribution. Here, μ_1 is the initial mean and μ_2 is the mean after a change has occurred according to model (2.2). The location of the changes are $(\mathcal{K}_1, \mathcal{K}_2) = (25\%, 25\% + 1), (25\%, 50\%), (25\%, 100\%)$.

[TABLE 2.9 HERE]

2.5 Behavior under serial dependence

One of the fundamental assumption for the use of the six test statistics is that X_1, \dots, X_n are independent. The aim of this subsection is to study the behavior of the tests when the series show serial dependence.

The privileged approach to create dependency in the series is based on copulas. The main reason is that copulas enable to generate random sequences with a chosen dependence structure and selected marginal distributions. If H is a bivariate distribution function with continuous margins F_1, F_2 , then there exists a unique copula \mathcal{C} such that $H(x_1, x_2) = \mathcal{C}\{F_1(x_1), F_2(x_2)\}$. Firstly, note that to generate a time series V_1, \dots, V_n of uniform variables for which $(V_i, V_{i+1}) \sim \mathcal{C}$ for all $1 \leq i < n$, one can use the following algorithm :

Algorithm 1

- (i) Generate U_1, \dots, U_n i.i.d. from the $\mathcal{U}(0, 1)$ distribution;
- (ii) Put $V_1 = U_1$ and for $i = 2, \dots, n$, let $V_i = F_{U_{i-1}}^{-1}(U_i)$, where

$$F_x(y) = \mathbb{P}(Y \leq y | X = x) = \frac{\partial}{\partial x} \mathcal{C}(x, y).$$

This procedure was used by Genest et al. (2002) in order to study the power of some tests of randomness. Among the most popular dependence structures, one has the copulas of Clayton (CL), Farlie–Gumbel–Morgenstern (FGM)

and Frank (F), whose expressions are, respectively,

$$\mathcal{C}_\theta^{\text{CL}}(u, v) = \max \left\{ (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, 0 \right\}, \quad \theta > -1,$$

$$\mathcal{C}_\theta^{\text{FGM}}(u, v) = uv + \theta uv(1-u)(1-v), \quad -1 \leq \theta \leq 1,$$

$$\mathcal{C}_\theta^{\text{F}}(u, v) = -\frac{1}{\theta} \ln \left\{ 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right\}, \quad \theta \neq 0.$$

The strength of the dependence induced by a given copula model can be measured by Kendall's tau, i.e.

$$\tau(\mathcal{C}) = 4 \int_0^1 \int_0^1 \mathcal{C}(u, v) d\mathcal{C}(u, v) - 1.$$

It is well known that $\tau(\mathcal{C}^{\text{CL}}) = \theta/(\theta + 2)$ and $\tau(\mathcal{C}^{\text{FGM}}) = 2\theta/9$, so that the dependence parameter θ can easily be assigned a value that matches a desired level for Kendall's tau. While Clayton and Frank families are complete in the sense that they can model all levels of dependence, the range of Kendall's tau for the Farlie-Gumbel-Morgenstern family is $[-2/9, 2/9]$.

Since $\tau(\mathcal{C}^{\text{F}})$ is not explicit, numerical methods must be used to compute $\tau^{-1}(\theta)$. In Table 2.10, the ability of the tests to keep their nominal level under series generated from *Algorithm 1* is studied. Note that in this model, the margins are arbitrarily taken to be standard uniform, since a change of scale would not affect the values of the ranks.

[TABLE 2.10 HERE]

In Table 2.11, the power of the tests to detect smooth changes in the mean is investigated under serial dependence. To this end, consider a location-scale distribution of the form

$$F_{\mu,\sigma}(x) = F\left(\frac{x - \mu}{\sigma}\right).$$

Here, σ is a fixed parameter, while the mean μ changes according to the smooth-change model (2.2). Consider the following algorithm :

Algorithm 2

- (i) *Generate a time series V_1, \dots, V_n , where $(V_i, V_{i+1}) \sim \mathcal{C}$;*
- (ii) *Let $X_i = \sigma F^{-1}(V_i) + \mu_i$, where*

$$\mu_i = \begin{cases} \mu_1, & 1 \leq i \leq \mathcal{K}_1; \\ \mu_1 + \frac{(i - \mathcal{K}_1)(\mu_2 - \mu_1)}{\mathcal{K}_2 - \mathcal{K}_1}, & \mathcal{K}_1 < i \leq \mathcal{K}_2; \\ \mu_2, & \mathcal{K}_2 < i \leq n. \end{cases}$$

[TABLE 2.11 HERE]

2.6 Discussion

Generally speaking, the test statistics S_{n1} and T_{n1} are powerful under scenarios of the type (\mathcal{M}_1) , S_{n2}, T_{n2} are powerful under scenarios of the type (\mathcal{M}_2) and S_{n3}, T_{n3} are powerful under scenarios of the type (\mathcal{M}_3) . However,

S_{n2} and T_{n2} are less powerful under (\mathcal{M}_3) , while S_{n3} and T_{n3} are less powerful under (\mathcal{M}_2) . The test statistics S_{n1} and T_{n1} have good power under the three considered situations. In most cases, the Cramér–von Mises statistics are more powerful than their Kolmogorov–Smirnov version.

The location of the breakpoint has an influence on the observed powers. It seems that the powers reach their maximum value when a change occur at the middle of the series. As expected, the power increases as the sample size and / or the departure from \mathcal{H}_0 increases for all six statistics that were investigated.

The tests for the detection of breakpoints in the mean are robust against heterogeneity of the variance. This property of robustness is unfortunately not shared by the tests of breakpoint detections in the variance. Under serial dependence, the tests have some problems in keeping their nominal level under \mathcal{H}_0 , even when the level of dependence is weak. This influence of dependence can also be seen on the observed powers of the tests of breakpoint detection in the mean.

Table 2.3: Power, estimated with the help of 10 000 replicates, of S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} to detect departures from \mathcal{H}_0 for i.i.d. sequences from the Normal distribution with unit variance, $n = 50$

Breakpoints ($\mathcal{K}_1, \mathcal{K}_2$)	Test statistics	$\mu_d = \mu_1 - \mu_0 $					
		.25	.5	.75	1	1.25	1.5
(25%, 25% + 1)	S_{n1}	7.83	17.88	34.53	55.31	72.99	87.01
	S_{n2}	7.07	15.48	28.13	45.09	61.09	75.82
	S_{n3}	8.20	19.22	38.15	61.72	80.40	92.57
	T_{n1}	8.04	19.01	36.48	57.88	75.52	88.58
	T_{n2}	7.99	18.97	36.07	57.15	74.59	87.73
	T_{n3}	8.09	19.14	38.87	64.25	83.71	95.16
(25%, 50%)	S_{n1}	9.93	26.02	50.25	75.14	91.45	97.69
	S_{n2}	9.26	22.33	43.64	67.74	85.60	94.98
	S_{n3}	10.20	27.11	51.87	77.36	92.86	98.35
	T_{n1}	10.29	26.00	50.03	74.20	90.97	97.41
	T_{n2}	10.19	25.78	49.62	73.69	90.36	97.20
	T_{n3}	9.88	26.33	51.77	76.85	92.94	98.33
(25%, 75%)	S_{n1}	9.71	26.34	52.10	76.48	91.90	97.96
	S_{n2}	9.51	25.50	50.01	74.60	90.49	97.31
	S_{n3}	9.79	26.57	52.19	76.70	91.99	97.98
	T_{n1}	9.82	26.19	51.03	75.68	91.31	97.60
	T_{n2}	9.86	26.18	50.94	75.50	91.21	97.50
	T_{n3}	9.44	24.74	48.85	73.39	89.90	97.09
(25%, 100%)	S_{n1}	7.81	18.32	36.94	59.05	78.37	90.97
	S_{n2}	7.95	18.84	37.76	60.16	79.39	91.51
	S_{n3}	7.92	18.46	37.06	59.01	78.21	90.94
	T_{n1}	8.03	18.65	37.04	59.17	78.22	90.89
	T_{n2}	8.03	18.63	37.20	59.31	78.41	91.02
	T_{n3}	8.08	17.49	34.37	54.31	72.80	87.25
(50%, 50% + 1)	S_{n1}	11.03	31.53	60.12	83.67	95.49	99.26
	S_{n2}	10.40	29.76	56.81	80.94	93.68	98.68
	S_{n3}	11.03	31.91	60.77	84.33	95.82	99.36
	T_{n1}	10.68	30.40	58.06	81.44	94.06	98.95
	T_{n2}	10.70	30.09	57.74	81.10	93.95	98.87
	T_{n3}	11.13	32.26	62.23	85.80	96.74	99.71
(50%, 100%)	S_{n1}	7.11	16.69	31.24	49.84	68.63	83.17
	S_{n2}	7.73	18.33	35.17	55.50	74.67	88.06
	S_{n3}	7.53	16.72	31.84	50.99	70.38	84.81
	T_{n1}	7.46	16.92	31.61	49.99	68.72	82.72
	T_{n2}	7.51	16.97	31.76	50.22	69.09	82.99
	T_{n3}	7.58	15.99	29.73	48.00	67.25	82.33

Table 2.4: Power, estimated with the help of 10 000 replicates, of S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} to detect departures from \mathcal{H}_0 for i.i.d. sequences from the Normal distribution with unit variance, $n = 100$

Breakpoints ($\mathcal{K}_1, \mathcal{K}_2$)	Tcst statistics	$\mu_d = \mu_1 - \mu_0 $					
		.25	.5	.75	1	1.25	1.5
(25%, 25% + 1)	S_{n1}	12.43	33.96	62.43	85.75	96.71	99.54
	S_{n2}	10.77	27.66	51.30	74.76	90.73	97.45
	S_{n3}	12.94	36.84	68.10	91.09	98.57	99.94
	T_{n1}	12.80	35.32	64.06	86.92	97.10	99.64
	T_{n2}	12.65	34.71	63.02	85.84	96.57	99.49
	T_{n3}	11.82	35.67	69.91	93.02	99.21	99.97
(25%, 50%)	S_{n1}	16.23	50.03	82.75	97.41	99.81	100.0
	S_{n2}	14.54	43.07	75.62	94.86	99.30	99.99
	S_{n3}	16.29	51.27	84.26	97.80	99.81	100.0
	T_{n1}	16.06	48.60	81.76	96.98	99.76	100.0
	T_{n2}	15.92	48.02	81.13	96.71	99.73	100.0
	T_{n3}	14.83	48.45	83.20	97.32	99.86	100.0
(25%, 75%)	S_{n1}	16.58	49.52	83.21	97.22	99.84	99.99
	S_{n2}	15.83	47.04	81.02	96.40	99.74	100.0
	S_{n3}	16.39	49.31	82.90	97.20	99.84	99.99
	T_{n1}	16.20	48.23	81.99	96.89	99.79	99.98
	T_{n2}	16.11	47.99	81.82	96.80	99.78	99.98
	T_{n3}	15.38	45.18	79.09	96.02	99.68	99.99
(25%, 100%)	S_{n1}	12.69	36.25	66.13	88.51	97.59	99.70
	S_{n2}	12.76	37.05	66.87	88.98	97.88	99.72
	S_{n3}	12.62	36.11	65.89	88.33	97.50	99.69
	T_{n1}	12.60	36.27	65.85	88.31	97.57	99.70
	T_{n2}	12.52	36.08	65.82	88.31	97.56	99.71
	T_{n3}	10.79	31.49	59.78	83.30	95.77	99.34
(50%, 50% + 1)	S_{n1}	19.36	58.77	90.43	99.16	99.96	100.0
	S_{n2}	18.27	55.07	87.66	98.60	99.91	100.0
	S_{n3}	19.43	58.91	90.56	99.20	99.98	100.0
	T_{n1}	18.35	55.74	88.41	98.88	99.91	100.0
	T_{n2}	18.09	55.46	88.20	98.81	99.91	100.0
	T_{n3}	18.50	58.19	91.35	99.34	99.97	100.0
(50%, 100%)	S_{n1}	11.39	29.92	57.27	81.01	94.34	98.94
	S_{n2}	12.47	32.70	62.68	86.04	96.54	99.56
	S_{n3}	11.46	30.06	58.74	82.91	95.37	99.31
	T_{n1}	11.61	29.71	57.36	80.97	94.17	98.92
	T_{n2}	11.60	29.80	57.52	81.07	94.23	98.95
	T_{n3}	10.43	27.02	53.99	79.22	93.89	98.87

Table 2.5: Power, estimated with the help of 10 000 replicates, of S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} to detect breakpoints in the variance for i.i.d. sequences of size $n = 50$ from the Normal distribution with zero mean

$(\mathcal{K}_1, \mathcal{K}_2)$	Test statistics	$\sigma_2/\sigma_1 = 1.5$	$\sigma_2/\sigma_1 = 2$	$\sigma_2/\sigma_1 = 2.5$
(25%, 25% + 1)	S_{n1}	13.24	29.75	42.90
	S_{n2}	11.96	23.73	33.87
	S_{n3}	14.18	32.35	49.37
	T_{n1}	14.50	31.32	46.30
	T_{n2}	14.43	30.67	45.39
	T_{n3}	13.32	31.67	51.50
(25%, 50%)	S_{n1}	21.47	52.37	71.87
	S_{n2}	19.12	44.78	62.51
	S_{n3}	22.34	54.89	75.93
	T_{n1}	21.54	52.19	72.11
	T_{n2}	21.24	51.38	71.33
	T_{n3}	21.39	53.22	75.15
(25%, 75%)	S_{n1}	25.52	56.41	78.48
	S_{n2}	24.92	52.96	74.07
	S_{n3}	25.56	56.82	79.01
	T_{n1}	25.19	55.98	77.88
	T_{n2}	24.96	55.57	77.31
	T_{n3}	23.38	52.64	75.21
(25%, 100%)	S_{n1}	18.85	45.90	69.82
	S_{n2}	19.94	46.03	68.44
	S_{n3}	19.17	45.66	69.68
	T_{n1}	19.45	45.98	69.88
	T_{n2}	19.36	45.79	69.76
	T_{n3}	17.53	41.15	64.47
(50%, 50% + 1)	S_{n1}	29.16	66.45	87.37
	S_{n2}	27.99	62.36	83.46
	S_{n3}	29.37	67.21	87.93
	T_{n1}	28.27	64.14	85.46
	T_{n2}	27.98	63.40	84.94
	T_{n3}	29.48	68.48	89.11
(50%, 100%)	S_{n1}	17.64	42.94	66.06
	S_{n2}	20.23	47.53	69.73
	S_{n3}	18.02	43.75	66.53
	T_{n1}	17.78	42.92	65.56
	T_{n2}	17.83	43.02	65.62
	T_{n3}	17.12	41.45	64.11

Table 2.6: Power, estimated with the help of 10 000 replicates, of S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} to detect breakpoints in the variance for i.i.d. sequences of size $n = 100$ from the Normal distribution with zero mean

$(\mathcal{K}_1, \mathcal{K}_2)$	Test statistics	$\sigma_2/\sigma_1 = 1.5$	$\sigma_2/\sigma_1 = 2$	$\sigma_2/\sigma_1 = 2.5$
(25%, 25% + 1)	S_{n1}	29.67	66.12	86.40
	S_{n2}	24.18	53.20	73.65
	S_{n3}	32.31	74.13	93.32
	T_{n1}	30.90	67.91	87.90
	T_{n2}	30.37	66.59	86.36
	T_{n3}	31.04	77.10	96.16
(25%, 50%)	S_{n1}	47.38	88.25	98.34
	S_{n2}	40.83	80.29	94.99
	S_{n3}	48.53	90.11	98.99
	T_{n1}	46.23	87.58	98.04
	T_{n2}	45.55	86.63	97.60
	T_{n3}	45.82	89.02	98.80
(25%, 75%)	S_{n1}	49.61	90.25	98.64
	S_{n2}	46.34	86.76	97.69
	S_{n3}	49.37	90.24	98.76
	T_{n1}	48.23	89.34	98.51
	T_{n2}	47.85	88.95	98.37
	T_{n3}	44.96	87.18	98.12
(25%, 100%)	S_{n1}	37.27	80.15	96.15
	S_{n2}	37.22	78.96	95.22
	S_{n3}	36.91	79.67	96.06
	T_{n1}	36.86	79.62	96.03
	T_{n2}	36.75	79.33	95.91
	T_{n3}	32.65	74.58	93.57
(50%, 50% + 1)	S_{n1}	57.32	94.94	99.64
	S_{n2}	53.42	92.78	99.26
	S_{n3}	57.26	95.23	99.71
	T_{n1}	54.10	93.46	99.36
	T_{n2}	53.72	93.15	99.28
	T_{n3}	56.94	95.86	99.84
(50%, 100%)	S_{n1}	33.02	76.31	93.35
	S_{n2}	36.62	79.61	94.59
	S_{n3}	33.72	77.13	93.78
	T_{n1}	32.86	75.61	93.00
	T_{n2}	32.71	75.57	92.96
	T_{n3}	30.47	73.21	92.38

Table 2.7: Power, estimated with the help of 10 000 replicates, of S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} to detect departures from \mathcal{H}_0 for i.i.d. sequences from the Exponential distribution. Initial mean : $\lambda_1 = 1$; sample size : $n = 50$

λ_2	Test statistics	Score function	$(\mathcal{K}_1, \mathcal{K}_2)$		
			(25%, 25% + 1)	(25%, 50%)	(25%, 100%)
1.5	S_{n1}	ϕ_1	17.52	11.20	22.05
		ϕ_2	6.08	5.86	7.33
	S_{n2}	ϕ_1	16.89	12.42	22.04
		ϕ_2	6.18	6.35	7.79
	S_{n3}	ϕ_1	17.85	11.35	21.67
		ϕ_2	6.13	5.90	7.55
	T_{n1}	ϕ_1	17.20	11.54	21.65
		ϕ_2	6.15	6.11	7.42
	T_{n2}	ϕ_1	17.11	11.53	21.55
		ϕ_2	6.07	6.13	7.36
	T_{n3}	ϕ_1	18.53	10.81	18.86
		ϕ_2	6.24	5.77	6.92
3	S_{n1}	ϕ_1	79.96	56.16	89.82
		ϕ_2	9.32	13.79	18.83
	S_{n2}	ϕ_1	76.53	59.98	88.78
		ϕ_2	9.33	16.63	21.12
	S_{n3}	ϕ_1	80.59	56.69	89.69
		ϕ_2	9.68	14.16	19.08
	T_{n1}	ϕ_1	77.96	55.80	89.54
		ϕ_2	9.28	14.44	18.72
	T_{n2}	ϕ_1	77.58	55.87	89.46
		ϕ_2	9.17	14.40	18.75
	T_{n3}	ϕ_1	82.13	54.64	86.10
		ϕ_2	9.41	13.08	16.92

Table 2.8: Size of the tests of breakpoint detection in the mean based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} under i.i.d. sequences from the Normal distribution with mean 0 and a smooth change in the variance.

(σ_1^2, σ_2^2)	Test statistics	$(\mathcal{K}_1, \mathcal{K}_2)$		
		(25%, 25% + 1)	(25%, 50%)	(25%, 100%)
(5, 1)	S_{n1}	8.51	7.51	4.88
	S_{n2}	5.94	5.53	4.09
	S_{n3}	8.63	7.44	4.77
	T_{n1}	8.94	7.49	4.90
	T_{n2}	8.89	7.56	4.90
	T_{n3}	6.53	7.43	3.37
(5, 5)	S_{n1}	4.68	4.72	4.64
	S_{n2}	4.13	4.41	3.98
	S_{n3}	4.43	4.66	4.48
	T_{n1}	4.70	4.89	4.78
	T_{n2}	4.73	4.91	4.79
	T_{n3}	3.26	3.05	3.19
(5, 10)	S_{n1}	3.79	5.30	4.84
	S_{n2}	3.95	5.18	4.07
	S_{n3}	3.82	5.29	4.78
	T_{n1}	3.51	5.22	4.86
	T_{n2}	3.61	5.24	4.85
	T_{n3}	3.85	4.81	3.37
(5, 1)	S_{n1}	8.82	7.39	5.02
	S_{n2}	6.46	5.81	4.81
	S_{n3}	9.17	7.51	4.94
	T_{n1}	9.29	7.42	5.16
	T_{n2}	9.31	7.53	5.19
	T_{n3}	7.60	8.83	4.50
(5, 5)	S_{n1}	5.26	4.93	4.94
	S_{n2}	4.80	4.56	4.58
	S_{n3}	5.30	4.79	4.72
	T_{n1}	5.58	5.07	4.90
	T_{n2}	5.55	5.05	4.94
	T_{n3}	4.44	4.07	3.98
(5, 10)	S_{n1}	4.23	5.93	5.42
	S_{n2}	4.81	6.20	5.24
	S_{n3}	4.42	5.93	5.42
	T_{n1}	4.19	5.78	5.72
	T_{n2}	4.24	5.86	5.81
	T_{n3}	4.90	5.90	4.59

Table 2.9: Size of the tests of breakpoint detection in the variance based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} under i.i.d. sequences from the Normal distribution with variance 1 and a smooth change in the mean, $n = 50$.

(μ_1, μ_2)	Test statistics	$(\mathcal{K}_1, \mathcal{K}_2)$		
		(25%, 25% + 1)	(25%, 50%)	(25%, 100%)
(0, 1)	S_{n1}	11.65	4.71	3.59
	S_{n2}	9.72	4.49	4.00
	S_{n3}	12.13	4.88	3.68
	T_{n1}	12.22	5.01	3.67
	T_{n2}	12.09	5.02	3.66
	T_{n3}	10.71	4.57	3.62
(0, 2)	S_{n1}	72.30	12.92	1.89
	S_{n2}	57.83	9.45	4.77
	S_{n3}	80.39	16.92	4.21
	T_{n1}	74.07	15.22	2.25
	T_{n2}	72.69	15.03	2.28
	T_{n3}	83.59	21.22	6.34
(0, 3)	S_{n1}	98.72	34.45	1.12
	S_{n2}	91.09	19.14	9.16
	S_{n3}	99.90	61.57	23.37
	T_{n1}	98.91	39.56	1.49
	T_{n2}	98.22	38.02	1.58
	T_{n3}	100.00	82.23	30.47

Table 2.10: Size of the tests of breakpoint detection in the mean based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} under serial dependence, $n = 100$.

Copula	Test statistics	r								
		.0	.025	.050	.075	.100	.125	.150	.175	.200
Clayton	S_{n1}	5.59	5.79	6.98	8.04	9.72	11.41	13.22	14.43	16.39
	S_{n2}	5.39	5.99	6.92	7.96	9.80	11.20	13.10	14.29	16.01
	S_{n3}	5.49	6.06	7.39	8.84	10.86	13.01	15.41	17.27	20.06
	T_{n1}	5.50	5.79	6.88	8.24	9.71	11.29	13.02	14.50	16.68
	T_{n2}	5.43	5.67	6.78	8.18	9.64	10.97	12.76	14.24	16.14
	T_{n3}	5.10	6.27	7.56	9.67	12.04	14.92	17.68	20.79	23.92
Frank	S_{n1}	5.17	6.38	6.88	8.31	9.29	10.92	12.13	14.27	15.21
	S_{n2}	5.01	6.32	6.75	8.13	9.09	10.59	11.91	13.85	14.81
	S_{n3}	5.12	6.48	7.27	9.02	10.43	12.18	14.14	16.91	18.48
	T_{n1}	5.25	6.48	6.97	8.25	9.35	10.64	12.12	14.37	15.23
	T_{n2}	5.17	6.41	6.93	8.18	9.19	10.55	11.95	14.01	15.01
	T_{n3}	5.02	6.23	7.57	9.76	12.22	14.27	16.63	19.99	22.58
F.-G.-M.	S_{n1}	5.02	5.89	7.43	7.73	9.68	11.02	12.92	13.43	15.84
	S_{n2}	5.09	5.54	7.33	7.59	9.29	10.94	12.60	12.99	15.03
	S_{n3}	4.94	5.94	7.81	8.24	10.62	12.34	14.64	15.69	18.82
	T_{n1}	5.07	5.77	7.43	7.95	9.79	10.81	12.71	13.33	15.81
	T_{n2}	4.98	5.70	7.33	7.76	9.60	10.66	12.42	13.01	15.44
	T_{n3}	4.57	5.93	8.13	9.11	11.39	14.12	17.25	19.08	22.42

Table 2.11: Power of the tests of breakpoint detection in the mean based on S_{n1} , S_{n2} , S_{n3} , T_{n1} , T_{n2} and T_{n3} under serial dependence based on Clayton and $|\mu_1 - \mu_2| = 0.25$, $n = 100$.

$(\mathcal{K}_1, \mathcal{K}_2)$	Test statistics	τ									
		.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
(25%, 25% + 1)	S_{n1}	71.72	68.42	66.33	66.24	66.03	69.03	73.63	78.82	85.53	92.69
	S_{n2}	59.82	58.59	57.37	58.42	59.95	64.17	69.39	74.82	79.10	83.87
	S_{n3}	77.69	75.53	75.23	78.25	81.23	86.27	91.63	95.62	98.05	99.40
	T_{n1}	73.46	69.87	68.10	68.15	68.30	70.98	76.17	81.49	88.12	94.12
	T_{n2}	72.27	68.54	66.16	65.79	65.41	67.78	72.53	77.25	82.74	88.78
	T_{n3}	79.15	78.78	79.53	82.62	86.01	90.62	94.89	97.63	98.92	99.75
(25%, 50%)	S_{n1}	90.10	86.26	83.26	79.69	77.35	76.90	79.16	83.01	88.64	94.69
	S_{n2}	84.52	80.01	77.60	74.40	73.03	73.27	76.05	79.73	85.04	91.45
	S_{n3}	91.26	88.21	86.35	85.32	86.19	88.92	92.86	96.18	98.32	99.51
	T_{n1}	89.09	85.34	82.62	79.60	77.55	77.52	80.55	84.60	89.98	95.68
	T_{n2}	88.53	84.64	81.44	77.90	75.48	74.97	77.32	81.03	86.62	92.85
	T_{n3}	90.29	88.62	87.67	87.98	89.61	92.07	95.36	97.56	99.03	99.73
(25%, 100%)	S_{n1}	74.78	73.53	70.84	69.35	68.29	70.38	74.13	79.97	85.99	93.15
	S_{n2}	75.32	73.86	71.24	69.59	68.92	70.08	72.56	78.00	83.23	90.53
	S_{n3}	74.45	74.65	73.79	74.76	77.52	82.60	88.96	93.62	97.33	99.25
	T_{n1}	74.26	73.08	70.73	69.55	68.99	71.40	75.68	81.69	87.77	94.23
	T_{n2}	74.15	72.95	70.60	69.24	68.27	69.83	73.02	78.52	84.35	91.62
	T_{n3}	68.83	72.18	74.47	78.02	81.90	87.28	92.60	95.97	98.38	99.56

CHAPITRE 3

MODÉLISATION DE PHÉNOMÈNES PAR LES COPULES

Au chapitre 2, il a été question de ruptures dans une série de données unidimensionnelles. En particulier, la puissance d'une procédure de détection de changements graduels de la moyenne et de la variance due à Lombard (1987) a été étudiée de façon exhaustive.

Dans les chapitres qui suivent, le cas multidimensionnel est traité. Plus spécifiquement, on développera une procédure statistique pour détecter des changements dans la structure de dépendance d'une série de données multidimensionnelles. Pour ce faire, on invoquera le théorème de Sklar (1959), qui permet d'extraire toute l'information au sujet de la dépendance dans un vecteur aléatoire à travers une fonction appelée *copule*. Notre approche se distingue donc des méthodes classiques de détection de changements dans la loi conjointe par la considération unique de la dépendance. Cette méthodologie permet donc d'identifier la source des ruptures, à savoir s'ils proviennent de changements dans les marges ou dans la dépendance.

Afin de familiariser le lecteur à la théorie et aux applications des copules, une

introduction est présentée dans ce chapitre. À la section 3.1, la définition d'une copule est d'abord donnée, suivie de l'énoncé de l'important résultat de Sklar (1959). Plusieurs exemples sont détaillés. La section 3.2 s'attarde aux mesures de dépendance, alors que la section 3.3 est dédiée à la modélisation de données multidimensionnelles par les copules. Une excellente introduction aux aspects théoriques des copules est offerte par Nelsen (2006).

3.1 Quelques considérations théoriques sur les copules

Une copule est une fonction de répartition multidimensionnelle dont les marges sont uniformes. Cette fonction permet de modéliser la structure de dépendance d'un vecteur aléatoire indépendamment des comportements individuels des variables. Dans la suite, on se concentrera sur le cas bidimensionnel afin de simplifier la notation. L'extension au cas général à p dimensions des résultats présentés dans ce chapitre est immédiate.

Une fonction $\mathcal{C} : [0, 1]^2 \rightarrow [0, 1]$ qui satisfait les deux conditions suivantes est appelée une copule.

$$(i) \quad \mathcal{C}(u_1, 0) = \mathcal{C}(0, u_2) = 0, \mathcal{C}(u_1, 1) = u_1 \text{ et } \mathcal{C}(1, u_2) = u_2;$$

$$(ii) \quad \text{Pour tout } u_1 \leq u_2 \text{ et pour tout } v_1 \leq v_2,$$

$$\mathcal{C}(u_2, v_2) - \mathcal{C}(u_2, v_1) - \mathcal{C}(u_1, v_2) + \mathcal{C}(u_1, v_1) \geq 0.$$

Les propriétés (i) et (ii) sont nécessaires et suffisantes pour assurer que \mathcal{C} soit une fonction de répartition dont les marges sont uniformes sur l'intervalle $[0, 1]$. On verra qu'il est possible, à l'aide des copules, de construire des distributions multidimensionnelles avec les marges désirées et une structure de dépendance donnée. L'étude de la dépendance entre des variables par les copules origine du théorème de Sklar (1959), énoncé ci-dessous.

Théorème 3.1. *Soit H , une fonction de répartition bidimensionnelle de fonctions de répartition marginales $F_1(x_1) = H(x_1, \infty)$ et $F_2(x_2) = H(\infty, x_2)$. Alors il existe une copule \mathcal{C} telle que pour tout $(x_1, x_2) \in \mathbb{R}^2$, on a*

$$H(x_1, x_2) = \mathcal{C} \{F_1(x_1), F_2(x_2)\}. \quad (3.1)$$

Si F_1 et F_2 sont continues, alors la copule \mathcal{C} est unique.

Ce résultat fait ressortir le fait qu'une fonction de répartition est constituée de ses marges et d'une fonction de dépendance qui lie les variables entre elles. On peut donc, à l'aide de ce résultat, construire une fonction de répartition avec une structure de dépendance donnée et les marges F_1 et F_2 désirées. Une illustration est présentée dans la suite.

Exemple 3.1. *La copule de Clayton est de la forme*

$$\mathcal{C}_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad (3.2)$$

où $\theta \geq 0$ est un paramètre de dépendance. Supposons que l'on désire construire un modèle bidimensionnel avec une structure de dépendance de type Clayton et des marges exponentielles, c'est-à-dire $F_1(x) = 1 - e^{-\lambda_1 x}$ et $F_2(x) =$

$1 - e^{-\lambda_2 x}$. Ce modèle est représenté par la fonction de répartition à trois paramètres

$$\begin{aligned} H_{\theta, \lambda_1, \lambda_2}(x, y) &= \mathcal{C}_\theta \{F_1(x), F_2(y)\} \\ &= \left\{ (1 - e^{-\lambda_1 x})^{-\theta} + (1 - e^{-\lambda_2 y})^{-\theta} - 1 \right\}^{-1/\theta}. \end{aligned}$$

La représentation (3.1) permet aussi d'extraire la copule d'une loi bidimensionnelle. En effet, par le changement de variables $u_1 = F_1(x_1)$ et $u_2 = F_2(x_2)$, on déduit

$$\mathcal{C}(u_1, u_2) = H \{F_1^{-1}(u_1), F_2^{-1}(u_2)\}.$$

L'exemple suivant en donne une illustration.

Exemple 3.2. Soit la fonction de répartition

$$H(x, y) = \sqrt{\frac{xy}{x + y - xy}}$$

définie sur $[0, 1]^2$. Bien qu'elles soient définies sur l'intervalle $[0, 1]$, les marges ne sont pas uniformes, et donc H n'est pas une copule. En fait, les lois marginales sont données par

$$F_1(x) = H(x, 1) = \sqrt{x} \quad \text{et} \quad F_2(y) = H(1, y) = \sqrt{y}.$$

Comme $F_1^{-1}(u) = u^2$ et $F_2^{-1}(v) = v^2$, on déduit que la copule de H est

$$\begin{aligned} \mathcal{C}(u, v) &= H(u^2, v^2) = \sqrt{\frac{u^2 v^2}{u^2 + v^2 - u^2 v^2}} \\ &= \left(\frac{u^2 + v^2 - u^2 v^2}{u^2 v^2} \right)^{-1/2} = (u^{-2} + v^{-2} - 1)^{-1/2}. \end{aligned}$$

On retrouve la copule de Clayton de l'équation (3.2) avec $\theta = 2$.

Une classe qui est fréquemment utilisée pour modéliser un vecteur aléatoire est la famille des copules Archimédiennes. Cette classe de modèles a été formellement introduite par Genest & MacKay (1986). Pour les définir, soit $\phi : [0, 1] \rightarrow [0, \infty)$, une fonction strictement décroissante et convexe qui satisfait $\phi(1) = 0$. Une copule bidimensionnelle \mathcal{C} est Archimédienne si elle admet la représentation

$$\mathcal{C}_\phi(u_1, u_2) = \phi^{-1} \{ \phi(u_1) + \phi(u_2) \}. \quad (3.3)$$

La fonction ϕ est appelée le *générateur* de \mathcal{C}_ϕ . À noter que ces copules se généralisent facilement au cas à p dimensions. En effet, une copule Archimédienne p -dimensionnelle est de la forme

$$\mathcal{C}_\phi(u_1, \dots, u_p) = \phi^{-1} \{ \phi(u_1) + \dots + \phi(u_p) \}.$$

Quelques conditions supplémentaires sur ϕ doivent cependant être émises afin d'assurer que \mathcal{C}_ϕ soit une copule p -dimensionnelle.

La copule de Clayton est un membre de la famille Archimédienne. En effet, avec le générateur

$$\phi_\theta(t) = \frac{t^{-\theta} - 1}{\theta},$$

on retrouve l'équation (3.2) par une application de (3.3). Dans le cas général, cette formule donne la copule de Clayton multidimensionnelle, à savoir

$$\mathcal{C}_\theta(u_1, \dots, u_p) = \left(\sum_{j=1}^p u_j^{-\theta} - d + 1 \right)^{-1/\theta}.$$

Le modèle de Ali-Mikhail-Haq est également un membre de la famille des

copules Archimédiennes. Son générateur est

$$\phi_\theta(t) = \frac{1}{1-\theta} \ln \left(\frac{1-\theta}{t} + \theta \right), \quad -1 \leq \theta \leq 1.$$

En dimension $p = 2$, cela induit la copule

$$C_\theta(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}, \quad -1 \leq \theta \leq 1.$$

Une copule particulière est celle associée à l'indépendance. Dans le cas à deux variables, elle est définie par $\Pi(u, v) = uv$. Cette copule est Archimédienne de générateur $\phi(t) = -\log t$. En effet, $\phi^{-1}(u) = e^{-u}$, ce qui amène

$$\begin{aligned} \Pi(u, v) &= \phi^{-1}\{\phi(u) + \phi(v)\} \\ &= \phi^{-1}(-\log u - \log v) = e^{-(\log uv)} = uv. \end{aligned}$$

On remarque aussi que Π est un cas particulier des copules de Clayton et de Ali–Mikhail–Haq lorsque $\theta \rightarrow 0$.

Il est possible de définir un intervalle à l'intérieur duquel toutes les copules se situent. Explicitement, on peut affirmer que toute copule \mathcal{C} satisfait, pour tout $(u_1, \dots, u_p) \in [0, 1]^p$,

$$W(u_1, u_2) \leq \mathcal{C}(u_1, u_2) \leq M(u_1, u_2),$$

où

$$W(u_1, u_2) = \max(u_1 + u_2 - 1, 0) \quad \text{et} \quad M(u_1, u_2) = \min(u_1, u_2).$$

Les fonctions W et M sont des copules. On les appelle les bornes inférieure et supérieure de Fréchet–Hoeffding.

Quand la copule \mathcal{C} d'un couple de variables aléatoires (X_1, X_2) est M , cela implique que X_2 est une fonction monotone croissante de X_1 . En effet, supposons que $X_2 = g(X_1)$, où g est croissante. Alors

$$\begin{aligned} H(x_1, x_2) &= \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) \\ &= \mathbb{P}(X_1 \leq x_1, g(X_1) \leq x_2) \\ &= \mathbb{P}(X_1 \leq x_1, X_1 \leq g^{-1}(x_2)) \\ &= \min\{x_1, g^{-1}(x_2)\} \\ &= M\{x_1, g^{-1}(x_2)\}. \end{aligned}$$

Comme la loi marginale de X_2 est $g^{-1}(x_2)$, on conclut immédiatement que la copule de H est M . De façon semblable, si X_2 est une fonction monotone décroissante de X_1 , on peut montrer que la copule de (X_1, X_2) est W .

Malgré leur popularité, les copules Archimédiennes sont néanmoins très contraignantes, surtout en plusieurs dimensions, car un seul générateur sert à décrire toute la dépendance entre plusieurs paires de variables.

3.2 Mesure et détection de la dépendance

Soit un échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$, c'est-à-dire n copies indépendantes d'un vecteur aléatoire (X, Y) . Il existe différentes façons d'analyser la dépendance entre X et Y . Souvent, on désire quantifier la force du lien de dépendance en calculant une mesure d'association. À ce titre, trois mesures sont très populaires, à savoir

- (i) le coefficient de corrélation de Pearson ;
- (ii) le coefficient de corrélation de rangs de Spearman ;
- (iii) le tau de Kendall.

Il est aussi possible d'évaluer la force d'un lien de dépendance par une méthode de détection graphique. Ces sujets sont élaborés dans les sous-sections qui suivent.

3.2.1 Mesures de dépendance

- (i) La statistique de Pearson est définie par

$$\rho_n = \frac{\sum_{i=1}^n (X_i - \bar{X}_n) (Y_i - \bar{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2 \sum_{i=1}^n (Y_i - \bar{Y}_n)^2}},$$

où \bar{X}_n et \bar{Y}_n sont les moyennes empiriques marginales. La statistique ρ_n permet d'estimer le coefficient de corrélation

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}},$$

où $\text{cov}(X, y)$ est la covariance entre X et Y et $\text{var}(X)$ est la variance de X . Ce coefficient est surtout apte à mesurer l'existence d'une relation linéaire entre les observations.

- (ii) Le rho de Spearman est une mesure d'association non-paramétrique basée sur les rangs des observations. Soient R_i , le rang de X_i parmi X_1, \dots, X_n

et S_i , le rang de Y_i parmi Y_1, \dots, Y_n . Le rho de Spearman est simplement défini comme la corrélation empirique mesurée sur la base des paires de rangs $(R_1, S_1), \dots, (R_n, S_n)$. Après quelques simplifications, on trouve

$$r_S = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (S_i - R_i)^2.$$

La statistique de rangs r_S permet d'estimer la version théorique du rho de Spearman, qui s'exprime en fonction de la copule d'une population par

$$\rho(\theta) = 12 \int_0^1 \int_0^1 \{C_\theta(u, v) - uv\} du dv. \quad (3.4)$$

Le rho de Spearman est souvent conseillé au détriment du coefficient de corrélation de Pearson car $E(r_S) = \pm 1$ si et seulement si les variables X et Y ont une dépendance fonctionnelle entre elles, c'est-à-dire qu'il existe une fonction ξ telle que $Y = \xi(X)$. On a vu que pour ρ_n , ce résultat est valide seulement pour $\xi(x) = Ax + B$. De plus, r_S est bien défini lorsque les observations sont de type ordinal.

(iii) Comme le rho de Spearman, le tau de Kendall est une mesure d'association non-paramétrique basée sur les rangs. Sa version théorique est

$$\tau(\theta) = 4 \int_0^1 \int_0^1 C_\theta(u, v) dC_\theta(u, v) - 1. \quad (3.5)$$

Le tau de Kendall empirique est défini par

$$\tau_n = \frac{P_n - Q_n}{\binom{n}{2}} = \frac{4P_n}{n(n-1)} - 1, \quad (3.6)$$

où P_n et Q_n représentent respectivement le nombre de paires concordantes et discordantes parmi l'échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$. On dit que deux paires de points (X_i, Y_i) et (X_j, Y_j) sont concordantes lorsque

$$(X_i - X_j)(Y_i - Y_j) > 0,$$

et discordantes quand

$$(X_i - X_j)(Y_i - Y_j) < 0.$$

La probabilité que l'événement $(X_i - X_j)(Y_i - Y_j) = 0$ se produise est nulle puisque les lois de X et Y sont supposées continues. Puisque le tau de Kendall est une mesure d'association non-paramétrique basée sur les rangs, il possède les mêmes avantages que r_S .

3.2.2 Détection graphique de la dépendance

Genest & Boies (2003) ont proposé une méthode graphique pour détecter la présence de dépendance. Pour bien la comprendre, la technique classique du *Khi-plot* sera d'abord expliquée.

Le Khi-plot

Inspirée des cartes de contrôle, cette méthode a été proposée pour la première fois par Fisher & Switzer (1985). Elle fut approfondie et illustrée par Fisher & Switzer (2001). Pour construire ce modèle, définissons d'abord pour chaque

$i \in \{1, \dots, n\}$,

$$\begin{aligned} H_i &= \frac{1}{n-1} \sum_{j \neq i} \mathbb{I}(X_j \leq X_i, Y_j \leq Y_i), \\ F_i &= \frac{1}{n-1} \sum_{j \neq i} \mathbb{I}(X_j \leq X_i), \\ G_i &= \frac{1}{n-1} \sum_{j \neq i} \mathbb{I}(Y_j \leq Y_i). \end{aligned}$$

À noter que $H_i, F_i, G_i, 1 \leq i \leq n$, peuvent s'écrire uniquement en fonction des rangs des observations. En effet, on a par exemple que

$$\mathbb{I}(X_j \leq X_i, Y_j \leq Y_i) = \mathbb{I}(R_j \leq R_i, S_j \leq S_i),$$

puisque $X_j \leq X_i, Y_j \leq Y_i$ si et seulement si $R_i \leq R_j, S_j \leq S_i$. La méthode du *Khi-plot* est donc entièrement fondée sur les rangs des observations.

Soient $H(x, y) = P(X \leq x, Y \leq y)$, $F(x) = P(X \leq x)$ et $G(y) = P(Y \leq y)$. L'hypothèse d'indépendance entre X et Y implique que $H(x, y) = F(x)G(y)$. Ainsi, on s'attend à ce que $H_i \approx F_i \times G_i$ sous l'indépendance. Définissons maintenant, pour $1 \leq i \leq n$,

$$\tilde{F}_i = F_i - \frac{1}{2} \quad \text{et} \quad \tilde{G}_i = G_i - \frac{1}{2}.$$

L'idée de Fisher & Switzer (1985) consiste à reporter sur un graphique les paires $(A_1, B_1), \dots, (A_n, B_n)$, où

$$A_i = 4 \operatorname{sign}(\tilde{F}_i \tilde{G}_i) \max(\tilde{F}_i^2, \tilde{G}_i^2)$$

et

$$B_i = \frac{H_i - (1/2 + \tilde{F}_i)(1/2 + \tilde{G}_i)}{\sqrt{(1/2 + \tilde{F}_i)(1/2 - \tilde{F}_i)(1/2 + \tilde{G}_i)(1/2 - \tilde{G}_i)}}.$$

Pour éviter l'influence des valeurs aberrantes, il est recommandé de ne considérer que les points tels que

$$|A_i| \leq 4 \left(\frac{1}{n-1} - \frac{1}{2} \right)^2.$$

Sous l'indépendance, la majorité des points (A_i, B_i) , $1 \leq i \leq n$, devraient se situer à l'intérieur d'une bande de confiance autour de l'axe horizontal. Fisher & Switzer (1985, 2001) suggèrent que les bornes de cette bande de confiance soient de la forme $\pm c_p/\sqrt{n}$, où c_p est tel que la probabilité de se situer à l'intérieur de la bande sous l'indépendance soit p . Pour $p = 0.9$, $p = 0.95$ et $p = 0.99$, on a respectivement $c_p = 1.54$, $c_p = 1.78$ et $c_p = 2.18$.

Le K-plot de Genest & Boies (2003)

Cet outil de visualisation de la dépendance basé sur les rangs a été récemment proposé par Genest & Boies (2003). Cette technique, inspirée de la droite de Henry, est basée sur H_1, \dots, H_n définis précédemment. Plus spécifiquement, on considère les statistiques d'ordre $H_{(1)}, \dots, H_{(n)}$, où

$$H_{(1)} \leq H_{(2)} \leq \dots \leq H_{(n)}.$$

On construit ensuite les paires

$$(W_{1:n}, H_{(1)}), \dots, (W_{n:n}, H_{(n)}),$$

où $W_{i:n}$ est l'espérance de la i -ème statistique d'ordre d'un échantillon de taille n tiré d'une population dont la fonction de répartition est

$$K(w) = w - w \ln w, \quad w \in [0, 1].$$

La fonction K correspond à la transformation intégrale de probabilité de la copule d'indépendance. En d'autres termes, $K(w) = P\{\Pi(U, V) \leq w\}$, où $(U, V) \sim \Pi$. En effet, comme $\Pi(u, v) = uv$, on a

$$\begin{aligned} K(w) = P(UV \leq w) &= \int_0^1 P(UV \leq w | V = v) dv \\ &= \int_0^1 P\left(U \leq \frac{w}{v}\right) dv \\ &= \int_0^w dv + \int_w^1 \frac{w}{v} dv = w - w \log w. \end{aligned}$$

Une majorité de points situés au-dessus de la diagonale $y = x$ indique la présence de dépendance positive, alors qu'une majorité de points situés sous cette diagonale suggère plutôt la présence d'une dépendance négative.

3.3 Modélisation de données multidimensionnelles par les copules

S'il existe une forte relation entre des variables aléatoires, il est peu recommandé de les modéliser séparément. En effet, on perdrait alors de l'information précieuse sur les liens d'interdépendance. Il est plus judicieux de modéliser conjointement ces variables par l'ajustement de lois à plusieurs variables. Parmi les possibilités, on retrouve les lois normale, gamma et elliptique multidimensionnelles. Une contrainte liée à l'utilisation de ces modèles est que les lois marginales sont imposées. L'utilisation des copules abolit cette contrainte et offre ainsi une plus grande flexibilité.

Dans ce qui suit, quelques étapes importantes pour modéliser des observations bidimensionnelles $(X_1, Y_1), \dots, (X_n, Y_n)$ par les copules sont décrites.

3.3.1 Estimation des paramètres d'une copule

Supposons qu'une famille de copules $\{\mathcal{C}_\theta; \theta \in \Theta\}$ est suggérée pour représenter la fonction de dépendance des variables aléatoires X et Y . Le paramètre θ étant supposé inconnu, il incombe d'avoir une méthode pour l'estimer efficacement. Pour ce faire, une des deux stratégies suivantes est habituellement employée : l'inversion d'une mesure d'association ou la méthode du pseudo maximum de vraisemblance.

Inversion d'une mesure d'association

Parmi les mesures d'association disponibles, on retrouve le rho de Spearman et le tau de Kendall. Leurs expressions théoriques en fonction du paramètre de dépendance θ d'une copule \mathcal{C}_θ , à savoir $\rho(\theta)$ et $\tau(\theta)$, sont présentées aux équations (3.4) et (3.5), respectivement. Pour estimer $\rho(\theta)$, on utilise le rho de Spearman empirique r_S . De façon semblable, on utilise le tau de Kendall empirique τ_n pour estimer $\tau(\theta)$.

Pour expliquer la méthode d'estimation par inversion, supposons que la copule de la paire aléatoire (X, Y) appartient à la famille de Farlie–Gumbel–Morgenstern. Dans ce cas, on a

$$\mathcal{C}_\theta(u, v) = uv + \theta uv(1 - u)(1 - v)$$

pour un certain $\theta \in [-1, 1]$. On peut alors montrer que

$$\tau(\theta) = \frac{2\theta}{9} \quad \text{et} \quad \rho(\theta) = \frac{\theta}{3}.$$

La résolution, en terme de θ , de l'équation $\tau(\theta) = \tau_n$, conduit à l'estimateur par inversion du tau de Kendall $\hat{\theta}_n = 9\tau_n/2$. De même, l'estimateur par inversion du rho de Spearman donne $\tilde{\theta}_n = 3r_S$. Cette stratégie d'estimation est donc une adaptation de la méthode classique des moments.

Pour développer une telle méthode dans un cadre plus général, soit un modèle de copules $\{\mathcal{C}_\theta; \theta \in \Theta\}$, où θ est unidimensionnel. Supposons de plus que la fonction $\tau(\theta)$ est une bijection continue dont la première dérivée existe. L'estimateur de θ par inversion du tau de Kendall est $\hat{\theta}_n = \tau^{-1}(\tau_n)$.

Pour établir le comportement asymptotique de $\hat{\theta}_n$, on note d'abord qu'un développement en série de Taylor d'ordre 1 autour de $\tau(\theta)$ donne

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{1}{\tau'(\theta)} \sqrt{n}\{\tau_n - \tau(\theta)\} + o_P(1).$$

Ensuite, d'un résultat classique que l'on peut retrouver dans Lee (1990), on déduit que $\sqrt{n}\{\tau_n - \tau(\theta)\}$ converge en loi vers une normale de moyenne 0 et de variance

$$\sigma_\tau^2(\theta) = 4\text{var}\{2\mathcal{C}_\theta(U, V) - U - V\},$$

où $(U, V) \sim \mathcal{C}_\theta$. Enfin, une application du théorème de Slutsky assure que $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en loi vers une normale de moyenne 0 et de variance $\sigma_\tau^2(\theta)/\{\tau'(\theta)\}^2$. Une étude sur l'efficacité de cette méthode d'estimation est présentée dans Genest et al. (2006).

On peut appliquer le même raisonnement à l'inversion du rho de Spearman. Cette fois, il s'agit de résoudre $\rho(\theta) = r_S$ pour déduire l'estimateur $\tilde{\theta}_n = \rho^{-1}(r_S)$, dont le comportement asymptotique sera déduit de celui de $\sqrt{n}\{r_S - \rho(\theta)\}$. À cet effet, voir Borkowf (2002).

Cette méthode d'inversion n'est pas toujours techniquement simple, puisqu'il arrive que les fonctions $\tau(\theta)$ et $\rho(\theta)$ ne soient pas explicitement inversibles. De plus, cette technique fonctionne seulement lorsque θ est unidimensionnel. La prochaine méthode, basée sur le maximum de vraisemblance, ne souffre pas de ces limitations.

Pseudo maximum de vraisemblance

La méthode du pseudo maximum de vraisemblance s'avère quelquefois plus efficace que les méthodes d'estimation par inversion. En outre, on peut l'utiliser pour l'estimation de paramètres multidimensionnels.

Pour décrire cette méthode, soit un couple (X, Y) dont la copule est \mathcal{C}_θ et les marges F et G . La fonction de répartition conjointe est donc $H(x, y) = \mathcal{C}_\theta\{F(x), G(y)\}$. La densité associée à H est

$$h_\theta(x, y) = c_\theta\{F(x), G(y)\} f(x)g(y),$$

où $c_\theta(u, v) = \partial^2 \mathcal{C}_\theta(u, v) / \partial u \partial v$ et f, g sont les densités de F, G respectivement. La fonction de vraisemblance s'écrit alors

$$L(\theta) = \prod_{i=1}^n c_\theta\{F(X_i), G(Y_i)\} f(X_i)g(Y_i).$$

Comme f et g ne dépendent pas de θ , on peut écrire

$$L(\theta) = K \prod_{i=1}^n c_{\theta} \{F(X_i), G(Y_i)\},$$

où K est un nombre qui ne dépend pas de θ . Cette écriture suppose cependant que les marges F et G sont connues, ce qui n'est habituellement pas le cas en pratique. Face à ce problème, Genest et al. (1995) ont proposé de remplacer F et G par leurs estimateurs naturels

$$F_n(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{I}(X_i \leq x) \quad \text{et} \quad G_n(y) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{I}(Y_i \leq y).$$

Les fonctions F_n et G_n s'appellent les fonctions de répartition empiriques. On travaille donc plutôt avec la fonction de pseudo vraisemblance

$$\tilde{L}(\theta) = \prod_{i=1}^n c_{\theta} \{F_n(X_i), G_n(Y_i)\}.$$

L'estimateur du pseudo maximum de vraisemblance (PMV) est

$$\hat{\theta}_{PMV} = \arg \max_{\theta \in \Theta} \tilde{L}(\theta).$$

La solution s'obtient généralement par des méthodes numériques. À noter que $\hat{\theta}_{PMV}$ est entièrement basé sur les rangs des observations, car $F_n(X_i) = R_i/(n+1)$ et $G_n(Y_i) = S_i/(n+1)$.

3.3.2 Tests d'adéquation

Il est naturel de se demander si un modèle de copules particulier explique bien la dépendance entre des variables aléatoires. Pour répondre à cette

question, on utilise les tests d'adéquation. Le développement de ces tests est relativement récent et ce domaine est toujours en pleine effervescence. Dans ce qui suit, une brève revue de littérature sur ces tests est présentée.

Genest & Rivest (1993) furent les premiers à introduire une procédure d'adéquation pour les copules. Ces chercheurs ont considéré la transformation intégrale de probabilité d'une copule paramétrique \mathcal{C}_θ , c'est-à-dire $K_\theta(t) = P\{C_\theta(U, V) \leq t\}$. Ces auteurs ont proposé son estimation par

$$K_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(W_{i,n} \leq t),$$

où

$$W_{i,n} = \frac{1}{n} \sum_{j=1}^n \mathbb{I}(X_j \leq X_i, Y_j \leq Y_i),$$

et ont ensuite proposé de comparer graphiquement K_n avec la fonction $K_{\hat{\theta}_n}$ qui correspond au modèle de copules supposé sous l'hypothèse nulle. Ici, $\hat{\theta}_n$ est un estimateur de θ .

Plus tard, Barbe et al. (1996) ont étudié le comportement asymptotique du *processus de Kendall*, à savoir la fonction empirique

$$\mathcal{K}_{n,\theta}(t) = \sqrt{n} \{K_n(t) - K_\theta(t)\}, \quad 0 \leq t \leq 1.$$

Il est bon de remarquer qu'ici, le paramètre θ est connu.

Wang & Wells (2000) ont exploité une version de \mathcal{K}_n avec un paramètre estimé pour tester l'adéquation à une famille de copules. Leur idée consiste à mesurer une distance entre K_n et $K_{\hat{\theta}_n}$. Ceci revient implicitement à calculer une fonctionnelle du processus

$$\mathcal{K}_n(t) = \sqrt{n} \{K_n(t) - K_{\hat{\theta}_n}(t)\}.$$

Ils ont appliqué leur procédure de sélection aux copules Archimédiennes dans le cas de données bidimensionnelles censurées, en basant leur test sur la statistique

$$S_{\xi,n} = \int_{\xi}^1 \{\mathcal{K}_n(t)\}^2 dt, \quad \xi > 0.$$

Puisque l'intégrale se calcule sur l'intervalle $[\xi, 1] \subset [0, 1]$, leur méthode est basée sur une version tronquée de \mathcal{K}_n . Un inconvénient majeur de la procédure de Wang & Wells (2000) est le choix arbitraire de ξ . De plus, Genest et al. (2006) démontrent que la méthode qu'ils décrivent pour obtenir la p -valeur de leur test n'est pas valide. Cela rend son utilisation dangereuse.

Pour palier ces problèmes, Genest et al. (2006) ont étudié le comportement asymptotique de la version non tronquée du processus de Kendall. Ils ont ensuite justifié formellement une méthode de ré-échantillonnage pour calculer une p -valeur. Les statistiques de test qu'ils ont proposées sont les distances de Cramér-von Mises et Kolmogorov-Smirnov, à savoir

$$S_n = \int_0^1 \{\mathcal{K}_n(t)\}^2 k_{\hat{\theta}_n}(t) dt \quad \text{et} \quad T_n = \sup_{0 \leq t \leq 1} |\mathcal{K}_n(t)|,$$

où $k_{\theta}(t)$ est la densité de $K_{\theta}(t)$. À noter aussi que ces tests ne sont pas restreints aux copules Archimédiennes et qu'ils peuvent s'employer en dimension plus grande que deux.

Il est utile de savoir que les statistiques de test S_n et T_n possèdent des expressions explicites qui facilitent leur usage. En effet, Genest et al. (2006)

ont démontré que

$$\begin{aligned} S_n = & \frac{n}{3} + n \sum_{j=1}^{n-1} K_n^2 \left(\frac{j}{n} \right) \left\{ K_{\hat{\theta}_n} \left(\frac{j+1}{n} \right) - K_{\hat{\theta}_n} \left(\frac{j}{n} \right) \right\} \\ & - n \sum_{j=1}^{n-1} K_n \left(\frac{j}{n} \right) \left\{ K_{\hat{\theta}_n}^2 \left(\frac{j+1}{n} \right) - K_{\hat{\theta}_n}^2 \left(\frac{j}{n} \right) \right\} \end{aligned}$$

et

$$T_n = \sqrt{n} \max_{i=0,1; 0 \leq j \leq n-1} \left\{ \left| K_n \left(\frac{j}{n} \right) - K_{\hat{\theta}_n} \left(\frac{j+1}{n} \right) \right| \right\}.$$

Tel que mentionné, les p -valeurs associées aux statistiques S_n et T_n se calculent par ré-échantillonnage. La méthode particulière proposée s'appelle le *bootstrap paramétrique*. En voici les étapes :

- (1) Estimer θ par un estimateur convergent $\hat{\theta}_n$;
- (2) Générer N échantillons aléatoires de taille n à partir de la loi $\mathcal{C}_{\hat{\theta}_n}$;
- (3) Pour chaque échantillon généré à l'étape (2), calculer la statistique de test afin d'obtenir N répliques de S_n , à savoir $S_n^{(1)}, \dots, S_n^{(N)}$;
- (4) Calculer la p -valeur du test, c'est-à-dire

$$p = \frac{1}{N} \sum_{j=1}^N (S_n^{(j)} > S_n).$$

De façon équivalente, on obtient la valeur critique de rejet de l'hypothèse nulle en considérant les statistiques d'ordre $S_{1:N} \leq \dots \leq S_{N:N}$. Au seuil α , on rejette l'hypothèse nulle lorsque $S_n > S_{(1-\alpha)N:N}$.

Pour les cas où K_θ n'a pas de forme explicite, Genest & Rémillard (2008) proposent un *bootstrap paramétrique* à deux niveaux pour calculer la p -valeur.

D'autres tests d'adéquation pour les copules ont été proposés récemment. Par exemple, Fermanian (2005) a élaboré un test basé sur l'estimation par noyau de la densité d'une copule. Également, Mesfioui et al. (2009) ont développé un test basé sur la fonction de dépendance de Spearman, à savoir $L_\theta(t) = P(UV \leq t)$, où $(U, V) \sim \mathcal{C}_\theta$, et son estimation par

$$L_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I} \left(\frac{R_i S_i}{n^2} \leq t \right).$$

Ils ont démontré la convergence faible du processus $\mathcal{L}_n(t) = \sqrt{n}\{L_n(t) - L_{\hat{\theta}_n}(t)\}$. Les tests basés sur les fonctionnelles de Cramér–von Mises et de Kolmogorov–Smirnov de \mathcal{L}_n se sont avérés efficaces dans de nombreuses situations. Il est suggéré au lecteur de consulter les articles de Genest et al. (2009) et de Berg (2009) pour des analyses exhaustives des puissances de plusieurs tests d'adéquation pour les copules.

Tel que vu dans ce chapitre, la théorie des copules fournit des outils très pratiques pour modéliser la dépendance dans un vecteur aléatoire. L'ajustement formel de copules à des données multidimensionnelles, par l'emploi de tests d'adéquation, a fait l'objet de nombreux travaux récents, dont Fermanian (2005), Genest et al. (2006), Genest et al. (2009), Berg (2009) et Mesfioui et al. (2009). La question de la détection de ruptures dans les copules a toutefois été très peu abordée dans la littérature. Ce sujet est abordé au chapitre suivant, où une nouvelle méthodologie est développée.

CHAPITRE 4

NOUVEAUX TESTS DE DÉTECTION D'UN POINT DE RUPTURE DANS LA DÉPENDANCE

4.1 Problématique

Des conditions externes peuvent influencer les paramètres d'une loi qui sert à modéliser un phénomène. Par exemple, les changements climatiques pourraient modifier les débits des rivières et les durées des crues. Une autre illustration peut être donnée en finance, où les fluctuations de la bourse peuvent modifier un modèle explicatif initial, ce qui peut influencer les rendements économiques attendus. Dans cette optique, il est intéressant de rechercher si un tel changement s'est produit dans les paramètres de la loi ou au niveau de la loi elle-même, et le cas échéant, à quel moment cette rupture a eu lieu. Cette recherche de points de rupture dans le modèle initial peut nous renseigner sur les possibles causes ayant initié ces changements.

La détection de points de rupture dans les lois univariées a été largement étudiée. On peut citer, entre autres, Pettitt (1979) et Lombard (1987). Voir

également le chapitre 2 de ce mémoire pour de plus amples détails. Le cas à plusieurs variables a aussi été traité, notamment par Csörgö & Horváth (1988) et Horváth et al. (1999). Les tests développés dans ces cas sont basés, implicitement ou explicitement, sur la fonction de répartition d'un vecteur aléatoire. Ces techniques reposent sur la modélisation multivariée classique des variables aléatoires où les comportements marginaux sont entièrement déterminés par le modèle choisi. Par exemple, dans le modèle de Student d -varié à ν degrés de liberté, les marges sont nécessairement Student à ν degrés de liberté. Cependant, l'usage de ces modèles a montré qu'ils sont souvent inaptes à expliquer correctement un phénomène.

Le but de ce chapitre est de développer une méthodologie permettant de détecter une rupture dans la copule sous-jacente à un phénomène. À notre connaissance, seuls Dias & Embrechts (2004) se sont intéressés à ce problème. Toutefois, leur procédure est critiquable à deux niveaux : (i) sa validité requiert que les marges soient connues, ce qui est rarement le cas en pratique et (ii) on doit supposer une famille paramétrique de copules. Pour ce dernier aspect, on constate que ce test permet de détecter une rupture à l'intérieur d'une famille supposée de copules. Ainsi, on doit faire l'hypothèse, par exemple, que la copule appartient à la famille de Clayton. Le test sert alors seulement à vérifier si le paramètre de la copule a changé. Cette hypothèse sur la forme de la copule est très forte et est rarement justifiable en pratique. Le test de Dias & Embrechts (2004) est décrit à la section 4.2

Une méthodologie beaucoup plus satisfaisante consiste à tester si la copule a changé en se gardant d'émettre des hypothèses sur la forme de la structure

de dépendance. Gombay & Horváth (1999) ont développé une procédure non-paramétrique qui, si correctement adaptée, permet de réaliser ce test. On supposerait cependant, implicitement, que les comportements marginaux sont stables, c'est-à-dire non sujets à des ruptures. Ce test est décrit à la section 4.3.

La section 4.4 décrit l'apport méthodologique principal de ce mémoire. Spécifiquement, un nouveau test de détection d'une rupture de copules est présenté. La validité du test requiert très peu d'hypothèses sur les marges. En fait, il est permis que les marges se modifient dans leur moyenne et / ou leur variance. On verra que le nouveau test est, d'une certaine façon, une généralisation de celui de Gombay & Horváth (1999). Il est également beaucoup plus général que celui de Dias & Embrechts (2004) car aucune hypothèse n'est émise concernant la forme de la structure de dépendance.

4.2 Test de Dias & Embrechts (2004)

Soient $\mathbf{X}_1, \dots, \mathbf{X}_n$, des vecteurs aléatoires indépendants à valeurs dans \mathbb{R}^m .

On suppose que les fonctions de répartition respectives sont

$$P(\mathbf{X}_i \leq \mathbf{x}) = F(\mathbf{x}; \theta_i, \eta_i), \quad i \in \{1, \dots, n\},$$

où $\theta_1, \dots, \theta_n$ et η_1, \dots, η_n sont des paramètres à valeurs dans \mathbb{R}^d et \mathbb{R}^p , respectivement. Les paramètres $\theta_1, \dots, \theta_n$ sont sujets à une possible rupture, tandis que η_1, \dots, η_n demeurent stables. Les hypothèses nulle et alternative

d'un point de rupture dans le premier paramètre s'énoncent comme suit :

$$\mathcal{H}_0 : \theta_1 = \theta_2 = \dots = \theta_n \quad \text{et} \quad \eta_1 = \eta_2 = \dots = \eta_n$$

vs

$$\mathcal{H}_1 : \theta_1 = \dots = \theta_K \neq \theta_{K+1} = \dots = \theta_n \quad \text{et} \quad \eta_1 = \eta_2 = \dots = \eta_n.$$

Ainsi, l'hypothèse alternative prévoit un changement pour le premier paramètre au point \mathcal{K} , alors que le deuxième paramètre demeure inchangé au cours du temps. On dit alors que \mathcal{K} est le *point de rupture*. Il est à noter que les paramètres sont supposés inconnus.

Dans le cas le plus simple, on suppose que le moment \mathcal{K} de la rupture potentielle est connu. Ainsi, le test du rapport des vraisemblances pour cette situation consiste à rejeter \mathcal{H}_0 pour de petites valeurs de la statistique

$$\Lambda_{\mathcal{K}} = \frac{\sup_{(\theta, \eta) \in \mathbb{R}^d \times \mathbb{R}^p} \prod_{i=1}^n f(X_i; \theta, \eta)}{\sup_{(\theta, \theta', \eta) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^p} \prod_{i=1}^{\mathcal{K}} f(X_i; \theta, \eta) \prod_{i=\mathcal{K}+1}^n f(X_i; \theta', \eta)}.$$

Toutes les conditions de régularité et d'efficacité nécessaires à la validité de la méthode du maximum de vraisemblance doivent être rencontrées. Pour plus de détails, voir par exemple Van der Vaart (1998). Posons

$$L_{\mathcal{K}}(\theta, \eta) = \sum_{i=1}^{\mathcal{K}} \log f(X_i; \theta, \eta) \quad \text{et} \quad L_{\mathcal{K}}^*(\theta, \eta) = \sum_{i=\mathcal{K}+1}^n \log f(X_i; \theta, \eta).$$

Le rapport des vraisemblances $\Lambda_{\mathcal{K}}$ peut alors s'écrire sous la forme

$$-2 \log \Lambda_{\mathcal{K}} = 2 \left\{ L_{\mathcal{K}}(\hat{\theta}_{\mathcal{K}}, \hat{\eta}_{\mathcal{K}}) + L_{\mathcal{K}}^*(\theta_{\mathcal{K}}^*, \eta_{\mathcal{K}}^*) - L_n(\hat{\theta}_n, \hat{\eta}_n) \right\},$$

où $(\hat{\theta}_{\mathcal{K}}, \hat{\eta}_{\mathcal{K}})$, $(\theta_{\mathcal{K}}^*, \eta_{\mathcal{K}}^*)$ et $(\hat{\theta}_n, \hat{\eta}_n)$ sont les estimateurs du maximum de vraisemblance respectivement pour $L_{\mathcal{K}}$, $L_{\mathcal{K}}^*$ et L_n . On rejette donc \mathcal{H}_0 lorsque $-2 \log \Lambda_{\mathcal{K}} > q_{\alpha}$, où q_{α} est tel que $P(\Lambda_{\mathcal{K}} > q_{\alpha} | \mathcal{H}_0) = \alpha$.

Puisque dans la majorité des situations, le temps de rupture \mathcal{K} n'est pas connu, Dias & Embrechts (2004) suggèrent de rejeter \mathcal{H}_0 pour des valeurs élevées de

$$Z_n = \max_{1 \leq \mathcal{K} < n} (-2 \log \Lambda_{\mathcal{K}}).$$

Comme on l'a déjà mentionné, cette approche possède quelques limitations. D'abord, cette méthode est paramétrique au sens où l'on doit supposer une forme précise, à des paramètres près, pour la loi des observations. Cette hypothèse est rarement justifiable en pratique. De plus, des auteurs mentionnent que leur méthodologie permet de tester une rupture dans la copule. Par contre, pour y arriver, ils doivent supposer que les lois marginales sont connues, ce qui représente une hypothèse extrêmement forte.

4.3 Test de Gombay & Horváth (1999)

Soient $(X_1, Y_1), \dots, (X_n, Y_n)$, des vecteurs aléatoires indépendants à valeurs dans \mathbb{R}^2 . On associe à ces couples aléatoires les fonctions de répartition conjointes $H_1(x, y), \dots, H_n(x, y)$, où $H_i(x, y) = P(X_i \leq x, Y_i \leq y)$, $1 \leq i \leq n$. Les hypothèses nulle et alternative d'un point de rupture dans la fonction de répartition s'énoncent alors par

$$\mathcal{H}_0 : H_1 = \dots = H_n \quad \text{vs} \quad \mathcal{H}_1 : H_1 = \dots = H_{\mathcal{K}} \neq H_{\mathcal{K}+1} = \dots = H_n,$$

où \mathcal{K} est le moment de la rupture. En général, on suppose que \mathcal{K} est inconnu.

L'idée de Gombay & Horváth (1999) consiste d'abord à séparer les observations en deux sous-échantillons $(X_1, Y_1), \dots, (X_{\mathcal{K}}, Y_{\mathcal{K}})$ et $(X_{\mathcal{K}+1}, Y_{\mathcal{K}+1}), \dots, (X_n, Y_n)$, où $1 < \mathcal{K} < n$. On calcule ensuite $\hat{T}_{\mathcal{K}}$ et $\hat{T}_{\mathcal{K}}^*$, à savoir les tau de Kendall empiriques basés respectivement sur le premier et le deuxième échantillon. Se référer à l'équation (3.6) pour l'estimation du tau de Kendall.

Dans le cas où la loi conjointe des observations change au point \mathcal{K} , on a, au vu de la relation entre le tau de Kendall théorique et la copule présentée à l'équation (3.5), que le tau de Kendall change au point \mathcal{K} . En effet, supposons que \mathcal{C} est la copule associée à une certaine loi conjointe H et que τ est le tau de Kendall associé à \mathcal{C} . De (3.5),

$$\begin{aligned} \tau &= 4 \int_0^1 \int_0^1 \mathcal{C}(u, v) d\mathcal{C}(u, v) - 1 \\ &= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{C}\{F(x), G(y)\} d\mathcal{C}\{F(x), G(y)\} - 1 \\ &= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x, y) dH(x, y) - 1, \end{aligned}$$

en vertu du Théorème 3.1 qui assure que $H(x, y) = \mathcal{C}\{F(x), G(y)\}$. Ainsi, dans le cas d'une rupture dans les fonctions de répartition H_1, \dots, H_n , on s'attend à ce que la différence $|\hat{T}_{\mathcal{K}} - \hat{T}_{\mathcal{K}}^*|$ soit grande. Comme le temps de rupture \mathcal{K} est supposé inconnu, il s'agit de considérer un maximum pris sur toutes les différences possibles, c'est-à-dire en faisant varier \mathcal{K} entre 2 et $n-1$.

De façon précise, la procédure de test de Gombay & Horváth (1999) consiste à rejeter \mathcal{H}_0 pour de grandes valeurs de la statistique

$$\hat{M}_n = \max_{1 < \mathcal{K} < n} \frac{\mathcal{K}(n - \mathcal{K})}{n^{3/2}} \left| \hat{T}_{\mathcal{K}} - \hat{T}_{\mathcal{K}}^* \right|, \quad (4.1)$$

où

$$\begin{aligned} \hat{T}_{\mathcal{K}} &= \frac{2}{\mathcal{K}(\mathcal{K} - 1)} \sum_{1 \leq i < j \leq \mathcal{K}} \mathbb{I}\{(X_i - X_j)(Y_i - Y_j) > 0\}, \\ \hat{T}_{\mathcal{K}}^* &= \frac{2}{(n - \mathcal{K})(n - \mathcal{K} - 1)} \sum_{\mathcal{K} < i < j \leq n} \mathbb{I}\{(X_i - X_j)(Y_i - Y_j) > 0\}. \end{aligned}$$

Le théorème suivant de Gombay & Horváth (1999) décrit la loi asymptotique de la statistique de test \hat{M}_n .

Théorème 4.2. *Sous \mathcal{H}_0 , la statistique \hat{M}_n converge en loi vers*

$$\sigma \sup_{0 \leq t \leq 1} |B(t)|$$

où σ dépend de la loi de $(X_1, Y_1), \dots, (X_n, Y_n)$ et B est un pont Brownien, c'est-à-dire que $\{B(t), t \in [0, 1]\}$ est un processus aléatoire continu de moyenne 0 et de fonction de covariance $\gamma(s, t) = \text{cov}\{B(s), B(t)\} = \min(s, t) - st$.

Selon le théorème 4.2, la statistique \hat{M}_n converge vers une loi qui s'écrit en fonction d'un certain paramètre σ qui dépend de la loi des observations. Comme la distribution des observations est supposée inconnue, cela pose un problème pour déterminer la valeur critique du test, qui dépend aussi de σ .

Pour contourner ce problème, Gombay & Horváth (1999) proposent d'imiter la loi inconnue de \hat{M}_n en utilisant une méthode de ré-échantillonnage. L'idée

générale est d'estimer un paramètre inconnu à partir de l'information contenue dans l'échantillon lui-même. Dans le présent contexte, posons d'abord, pour $h \in \{1, \dots, N\}$,

$$\begin{aligned} U_n^{(h)}(\mathcal{K}) &= 2n^{1/2} \left[\frac{n - \mathcal{K}}{\mathcal{K}n} \sum_{1 \leq i, j \leq \mathcal{K}} \left(\epsilon_i^{(h)} - \hat{\epsilon}_{\mathcal{K}}^{(h)} \right) \mathbb{I} \{ (X_i - X_j)(Y_i - Y_j) > 0 \} \right. \\ &\quad \left. - \frac{\mathcal{K}}{n(n - \mathcal{K})} \sum_{\mathcal{K} < i, j \leq n} \left(\epsilon_i^{(h)} - \hat{\epsilon}_{\mathcal{K}}^{\star, (h)} \right) \mathbb{I} \{ (X_i - X_j)(Y_i - Y_j) > 0 \} \right], \end{aligned}$$

où $\epsilon^{(h)} = (\epsilon_1^{(h)}, \dots, \epsilon_n^{(h)})$ est un vecteur de variables indépendantes de moyenne nulle et de variance 1,

$$\hat{\epsilon}_{\mathcal{K}}^{(h)} = \frac{1}{\mathcal{K}} \sum_{1 \leq i \leq \mathcal{K}} \epsilon_i^{(h)} \quad \text{et} \quad \hat{\epsilon}_{\mathcal{K}}^{\star, (h)} = \frac{1}{n - \mathcal{K}} \sum_{\mathcal{K} < i \leq n} \epsilon_i^{(h)}.$$

Ensuite, on considère $M_n^{(1)}, \dots, M_n^{(N)}$, où

$$M_n^{(h)} = \max_{1 < \mathcal{K} < n} |U_n^{(h)}(\mathcal{K})|.$$

Gombay & Horváth (1999) démontrent que les valeurs $M_n^{(1)}, \dots, M_n^{(N)}$ sont asymptotiquement indépendantes et convergent en loi vers la même limite que \hat{M}_n , c'est-à-dire vers $\sigma \sup_{0 \leq t \leq 1} |B(t)|$. Ainsi, une valeur critique approximative pour un test de niveau α basé sur \hat{M}_n est donnée par le $100 \times (1 - \alpha)$ -ème percentile de $M_n^{(1)}, \dots, M_n^{(N)}$. Pour n et N suffisamment grands, on obtient donc une procédure de test valide.

À noter que si on suppose que les comportements marginaux ne sont pas sujets à des ruptures, le test basé sur \hat{M}_n permet, implicitement, de tester

une rupture dans la copule. En effet, supposons que $H_1 = \dots = H_K \neq H_{K+1} = \dots = H_n$. Du théorème de Sklar (1959), on sait qu'il existe des copules $\mathcal{C}_1, \dots, \mathcal{C}_n$ telles que $H_i(x, y) = \mathcal{C}_i\{F_i(x), G_i(y)\}$, $1 \leq i \leq n$. En supposant que $F_1 = \dots = F_n = F$ et $G_1 = \dots = G_n = G$, c'est-à-dire que les marges sont constantes, l'hypothèse de rupture s'écrit alors $\mathcal{C}_1\{F(x), G(y)\} = \dots = \mathcal{C}_K\{F(x), G(y)\} \neq \mathcal{C}_{K+1}\{F(x), G(y)\} = \dots = \mathcal{C}_n\{F(x), G(y)\}$. Ceci est équivalent à $\mathcal{C}_1 = \dots = \mathcal{C}_K \neq \mathcal{C}_{K+1} = \dots = \mathcal{C}_n$. Cependant, cette hypothèse sur la stabilité des marges n'est pas réaliste pour la plupart des applications. Pour cette raison, une modification du test de Gombay & Horváth (1999) est proposée à la section suivante.

4.4 Description d'une nouvelle méthodologie

4.4.1 Contexte

Dans ce qui suit, un nouveau test de détection non-paramétrique inspiré de celui de Gombay & Horváth (1999) est proposé. L'innovation principale est qu'il permet aux marges de se modifier dans leur moyenne et / ou variance au cours du temps. Pour ce faire, soit la série d'observations

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}.$$

Pour suivre la notation de la section précédente, définissons par \mathcal{C}_i la copule sous-jacente à la loi de (X_i, Y_i) et F_i, G_i , les marges. Ainsi, du théorème de

Sklar (1959), la fonction de répartition du couple (X_i, Y_i) est

$$H_i(x, y) = \mathcal{C}_i \{F_i(x), G_i(y)\}, \quad 1 \leq i \leq n.$$

On souhaite vérifier si la copule est demeurée la même au cours du temps, ou au contraire, s'il y a un certain temps $1 < \mathcal{K} < n$ à partir duquel la loi de dépendance change. Autrement dit, on désire confronter les hypothèses

$$\mathcal{H}_0 : \mathcal{C}_1 = \cdots = \mathcal{C}_{\mathcal{K}} = \mathcal{C}_{\mathcal{K}+1} = \cdots = \mathcal{C}_n \quad \text{vs}$$

$$\mathcal{H}_1 : \mathcal{C}_1 = \cdots = \mathcal{C}_{\mathcal{K}} \neq \mathcal{C}_{\mathcal{K}+1} = \cdots = \mathcal{C}_n \quad \text{pour un certain } \mathcal{K} \in \{1, \dots, n\}.$$

Il s'agit donc d'hypothèses sur un changement abrupt de copules.

On rappelle que dans la méthodologie développée par Gombay & Horváth (1999), il est supposé implicitement que les lois marginales sont homogènes au cours du temps, c'est-à-dire que $X_i \sim F$ et $Y_i \sim G$ pour tout $1 \leq i \leq n$. Dans ce contexte, l'utilisation des copules devient inutile car la détection d'un changement dans la dépendance revient à détecter un changement dans la loi conjointe $H(x, y) = P(X \leq x, Y \leq y)$. On se retrouve, en fait, dans un schéma qui a été abondamment étudié dans la littérature. Pour détecter une rupture dans la dépendance sans que cela ne soit éventuellement affecté par des ruptures dans les lois marginales, il faut définir des estimateurs adéquats pour le tau de Kendall, qui sert de base au test. Sinon, les changements éventuels dans les marges rendront les estimateurs $\hat{T}_{\mathcal{K}}$ et $\hat{T}_{\mathcal{K}}^*$ biaisés.

4.4.2 Stabilisation des marges : en théorie

Pour éviter le problème des marges qui peuvent se modifier au cours du temps, supposons que les moyennes et les variances de X et Y sont sujettes à d'éventuels changements selon le modèle de rupture graduelle de Lombard (1987). De plus, pour faire en sorte que ces changements caractérisent complètement les comportements de X et de Y , il est supposé que les lois marginales appartiennent à une famille de localisation lieu-échelle. Autrement dit, on suppose qu'il existe des fonctions de répartition F et G telles que

$$F_i(x) = F\left(\frac{x - \mu_i}{\sigma_i}\right) \quad \text{et} \quad G_i(y) = G\left(\frac{y - \nu_i}{\gamma_i}\right).$$

Plus précisément, supposons qu'entre les temps \mathcal{K}_{μ_1} et \mathcal{K}_{μ_2} , il y a eu un changement graduel de moyenne et qu'entre les temps \mathcal{K}_{σ_1} et \mathcal{K}_{σ_2} , il y a eu un changement graduel d'écart-type dans la loi de X . De même, soient les temps \mathcal{K}_{ν_1} , \mathcal{K}_{ν_2} , \mathcal{K}_{γ_1} et \mathcal{K}_{γ_2} où des changements graduels de moyenne et de variance se sont produits dans la loi de Y . Selon le modèle de Lombard (1987), on a

$$\mu_i = \mu_1 + \frac{(i - \mathcal{K}_{\mu_1})(\mu_2 - \mu_1)}{\mathcal{K}_{\mu_2} - \mathcal{K}_{\mu_1}}, \quad \mathcal{K}_{\mu_1} < i \leq \mathcal{K}_{\mu_2}$$

et

$$\sigma_i = \sigma_1 + \frac{(i - \mathcal{K}_{\sigma_1})(\sigma_2 - \sigma_1)}{\mathcal{K}_{\sigma_2} - \mathcal{K}_{\sigma_1}}, \quad \mathcal{K}_{\sigma_1} < i \leq \mathcal{K}_{\sigma_2}.$$

On peut ainsi décrire la fonction de répartition de X en fonction de ces paramètres qui se modifient au fil du temps. Par exemple, si on suppose que

$\mathcal{K}_{\mu_1} < \mathcal{K}_{\mu_2} < \mathcal{K}_{\sigma_1} < \mathcal{K}_{\sigma_2}$, alors

$$F_i(x) = P(X_i \leq x) = \begin{cases} F\left(\frac{x - \mu_1}{\sigma_1}\right), & 1 \leq i \leq \mathcal{K}_{\mu_1}; \\ F\left(\frac{x - \mu_i}{\sigma_1}\right), & \mathcal{K}_{\mu_1} < i \leq \mathcal{K}_{\mu_2}; \\ F\left(\frac{x - \mu_2}{\sigma_1}\right), & \mathcal{K}_{\mu_2} < i \leq \mathcal{K}_{\sigma_1}; \\ F\left(\frac{x - \mu_2}{\sigma_i}\right), & \mathcal{K}_{\sigma_1} < i \leq \mathcal{K}_{\sigma_2}; \\ F\left(\frac{x - \mu_2}{\sigma_2}\right), & \mathcal{K}_{\sigma_2} < i \leq n. \end{cases}$$

Pour faire en sorte que la moyenne de X soit constante dans le temps, on construit l'échantillon transformé $\tilde{X}_1, \dots, \tilde{X}_n$, où

$$\tilde{X}_i = \begin{cases} X_i, & i \leq \mathcal{K}_{\mu_1}; \\ X_i - \frac{(i - \mathcal{K}_{\mu_1})(\mu_2 - \mu_1)}{\mathcal{K}_{\mu_2} - \mathcal{K}_{\mu_1}}, & \mathcal{K}_{\mu_1} < i \leq \mathcal{K}_{\mu_2}; \\ X_i - (\mu_2 - \mu_1), & i > \mathcal{K}_{\mu_2}. \end{cases}$$

On a alors $E(\tilde{X}_i) = \mu_1$ pour tout $i \in \{1, \dots, n\}$. Ensuite, puisque $\text{var}(\tilde{X}_i) = \text{var}(X_i)$, on rend l'écart-type constant en supposant un échantillon $\tilde{\tilde{X}}_1, \dots, \tilde{\tilde{X}}_n$, où $\tilde{\tilde{X}}_i = a_i \tilde{X}_i + b_i$. Les constantes a_1, \dots, a_n et b_1, \dots, b_n sont choisies de telle

sorte que $E(\tilde{X}_i) = \mu_1$ et $\sqrt{\text{var}(\tilde{X}_i)} = \sigma_1$. On trouve

$$(a_i, b_i) = \begin{cases} (1, 0), & i \leq \mathcal{K}_{\sigma_1}; \\ \frac{((\mathcal{K}_{\sigma_2} - \mathcal{K}_{\sigma_1})\sigma_1, (i - \mathcal{K}_{\sigma_1})(\sigma_2 - \sigma_1)\mu_1)}{(\mathcal{K}_{\sigma_2} - \mathcal{K}_{\sigma_1})\sigma_1 + (i - \mathcal{K}_{\sigma_1})(\sigma_2 - \sigma_1)}, & \mathcal{K}_{\sigma_1} < i \leq \mathcal{K}_{\sigma_2}; \\ \left(\frac{\sigma_1}{\sigma_2}, \frac{\mu_1(\sigma_2 - \sigma_1)}{\sigma_2} \right), & i > \mathcal{K}_{\sigma_2}. \end{cases}$$

De même, supposons que la moyenne et la variance de Y_1, \dots, Y_n sont

$$\nu_i = \nu_1 + \frac{(i - \mathcal{K}_{\nu_1})(\nu_2 - \nu_1)}{\mathcal{K}_{\nu_2} - \mathcal{K}_{\nu_1}}, \quad \mathcal{K}_{\nu_1} < i \leq \mathcal{K}_{\nu_2}$$

et

$$\gamma_i = \gamma_1 + \frac{(i - \mathcal{K}_{\gamma_1})(\gamma_2 - \gamma_1)}{\mathcal{K}_{\gamma_2} - \mathcal{K}_{\gamma_1}}, \quad \mathcal{K}_{\gamma_1} < i \leq \mathcal{K}_{\gamma_2}.$$

Pour que la moyenne soit constante, on construit l'échantillon transformé $\tilde{Y}_1, \dots, \tilde{Y}_n$, où

$$\tilde{Y}_i = \begin{cases} Y_i, & i \leq \mathcal{K}_{\nu_1}; \\ Y_i - \frac{(i - \mathcal{K}_{\nu_1})(\nu_2 - \nu_1)}{\mathcal{K}_{\nu_2} - \mathcal{K}_{\nu_1}}, & \mathcal{K}_{\nu_1} < i \leq \mathcal{K}_{\nu_2}; \\ X_i - (\nu_2 - \nu_1), & i > \mathcal{K}_{\nu_2}. \end{cases}$$

On a alors $E(\tilde{Y}_i) = \nu_1$ pour tout $i \in \{1, \dots, n\}$. On rend ensuite l'écart-type constant en supposant un échantillon $\tilde{\tilde{Y}}_1, \dots, \tilde{\tilde{Y}}_n$, où $\tilde{\tilde{Y}}_i = c_i \tilde{Y}_i + d_i$. Les constantes c_1, \dots, c_n et d_1, \dots, d_n sont choisies de telle sorte que $E(\tilde{\tilde{Y}}_i) = \nu_1$

et $\sqrt{\text{var}(\tilde{Y}_i)} = \gamma_1$. On trouve alors

$$(c_i, d_i) = \begin{cases} (1, 0), & i \leq \mathcal{K}_{\gamma_1}; \\ \frac{((\mathcal{K}_{\gamma_2} - \mathcal{K}_{\gamma_1})\gamma_1, (i - \mathcal{K}_{\gamma_1})(\gamma_2 - \gamma_1)\nu_1)}{(\mathcal{K}_{\gamma_2} - \mathcal{K}_{\gamma_1})\gamma_1 + (i - \mathcal{K}_{\gamma_1})(\gamma_2 - \gamma_1)}, & \mathcal{K}_{\gamma_1} < i \leq \mathcal{K}_{\gamma_2}; \\ \left(\frac{\gamma_1}{\gamma_2}, \frac{\nu_1(\gamma_2 - \gamma_1)}{\gamma_2} \right), & i > \mathcal{K}_{\gamma_2}. \end{cases}$$

4.4.3 Stabilisation des marges : en pratique

En pratique, on ne connaît pas les valeurs $\mu_1, \mu_2, \sigma_1, \sigma_2, \nu_1, \nu_2, \gamma_1$ et γ_2 . Une façon simple consiste à les estimer par leurs estimateurs naturels, à savoir les moyennes et les écart-types empiriques. Ainsi, supposons que l'on sait que des changements de moyennes et d'écart-types sont survenus aux temps $\mathcal{K}_{\mu_1}, \mathcal{K}_{\mu_2}, \mathcal{K}_{\sigma_1}$ et \mathcal{K}_{σ_2} pour la loi de X . À la sous-section suivante, on verra comment identifier ces points de rupture. On définit alors

$$\begin{aligned} \hat{\mu}_1 &= \frac{1}{\mathcal{K}_{\mu_1}} \sum_{i=1}^{\mathcal{K}_{\mu_1}} X_i, \\ \hat{\mu}_2 &= \frac{1}{n+1-\mathcal{K}_{\mu_2}} \sum_{i=\mathcal{K}_{\mu_2}}^n X_i, \\ \hat{\sigma}_1^2 &= \frac{1}{\mathcal{K}_{\sigma_1}-1} \sum_{i=1}^{\mathcal{K}_{\sigma_1}} (X_i - \hat{\mu}_1)^2 \\ \hat{\sigma}_2^2 &= \frac{1}{n-\mathcal{K}_{\sigma_2}} \sum_{i=\mathcal{K}_{\sigma_2}}^n (X_i - \hat{\mu}_2)^2. \end{aligned}$$

On applique ensuite les transformations identifiées à la sous-section précédente

en remplaçant les valeurs des paramètres par leurs estimations.

Pour stabiliser la loi de Y_1, \dots, Y_n , on procède de façon similaire. On définit

$$\begin{aligned}\hat{\nu}_1 &= \frac{1}{\mathcal{K}_{\nu_1}} \sum_{i=1}^{\mathcal{K}_{\nu_1}} Y_i, \\ \hat{\nu}_2 &= \frac{1}{n+1-\mathcal{K}_{\nu_2}} \sum_{i=\mathcal{K}_{\nu_2}}^n Y_i, \\ \hat{\gamma}_1^2 &= \frac{1}{\mathcal{K}_{\gamma_1}-1} \sum_{i=1}^{\mathcal{K}_{\gamma_1}} (Y_i - \hat{\nu}_1)^2 \\ \hat{\gamma}_2^2 &= \frac{1}{n-\mathcal{K}_{\gamma_2}} \sum_{i=\mathcal{K}_{\gamma_2}}^n (Y_i - \hat{\nu}_2)^2,\end{aligned}$$

et on remplace ces estimations dans les transformations de la sous-section précédente.

4.4.4 Identification des points de rupture marginaux

La méthode de transformation décrite précédemment nécessite une procédure statistique qui permet de détecter les possibles ruptures de moyennes et de variances dans une série chronologique univariée. Une telle procédure est décrite par Lombard (1987), en marge des procédures qu'il a développées pour identifier des changements graduels. Pour de plus amples détails et des résultats de simulations, voir le chapitre 2.

Dans un cas où un changement de moyenne ou de variance est identifié par la procédure de Lombard (1987), on estime les temps de rupture par $(\hat{\mathcal{K}}_1, \hat{\mathcal{K}}_2)$

de telle sorte que le rapport

$$\frac{|L_{\mathcal{K}_1, \mathcal{K}_2}|}{\sigma\left(\frac{\mathcal{K}_1}{n}, \frac{\mathcal{K}_2}{n}\right)}$$

soit maximisé, où $L_{\mathcal{K}_1, \mathcal{K}_2}$ est définie à l'équation (2.4) et

$$\sigma^2(u, v) = \frac{(1-u)^3(1+3u)}{12} - \frac{(1-v)^3(1+3v)}{12} - \frac{(1-v)^2(v^2-u^2)}{2}.$$

Ces estimations permettent ensuite d'homogénéiser la moyenne et la variance des séries X_1, \dots, X_n et Y_1, \dots, Y_n . Une fois que les séries sont stabilisées, il est possible d'appliquer le test de Gombay & Horváth (1999), puisque celui-ci nécessite que les marges soient constantes. À noter qu'ici, les marges sont *pseudo-constantes*, puisque les transformations ne sont pas exactes. En effet, elles dépendent d'abord de la conclusion des tests de détection, avec probabilité d'erreur α , ainsi que des estimations de moyennes et de variances.

4.4.5 Étapes de conduite du test

Cette section présente les trois étapes à effectuer pour procéder à notre nouveau test de détection d'une rupture de copules. Au vu des résultats de puissance et de robustesse présentés au chapitre 2, la statistique qui sera utilisée pour la détection d'une rupture sera S_{n1} . Cette statistique s'est avérée extrêmement efficace pour détecter des changements graduels de moyenne et de variance, pour lesquels elle est conçue, mais aussi des changements abrupts.

Étape 1. Stabilisation de la série X_1, \dots, X_n

- (i) Vérifier s'il y a une rupture de moyenne dans la série X_1, \dots, X_n à l'aide de la statistique S_{n1} de Lombard (1987) avec le score $\phi_1(u) = 2u - 1$
- (ii) S'il y a effectivement une modification de la moyenne de la série, identifier les temps $\hat{\mathcal{K}}_{\mu_1}$ et $\hat{\mathcal{K}}_{\mu_2}$ où celle-ci s'est produite
- (iii) Transformer la série X_1, \dots, X_n en $\tilde{X}_1, \dots, \tilde{X}_n$
- (iv) Vérifier s'il y a une rupture de variance dans la série $\tilde{X}_1, \dots, \tilde{X}_n$ à l'aide de la statistique S_{n1} de Lombard (1987) avec les scores $\phi_2(u) = (2u-1)^2$ ou $\phi(u) = \log(1-u)$
- (v) S'il y a effectivement une modification de la variance de la série, identifier les temps $\hat{\mathcal{K}}_{\sigma_1}$ et $\hat{\mathcal{K}}_{\sigma_2}$ où celle-ci s'est produite
- (vi) Transformer la série $\tilde{X}_1, \dots, \tilde{X}_n$ en $\tilde{\tilde{X}}_1, \dots, \tilde{\tilde{X}}_n$

Étape 2. Stabilisation de la série Y_1, \dots, Y_n

Suivre une démarche similaire à l'étape 1.

Étape 3. Appliquer le test de Gombay & Horváth (1999) à la série bivariable transformée $(\tilde{\tilde{X}}_1, \tilde{\tilde{Y}}_1), \dots, (\tilde{\tilde{X}}_n, \tilde{\tilde{Y}}_n)$. Conclure quant à la rupture ou non dans la loi de dépendance, c'est-à-dire la copule. S'il y a effectivement une rupture, identifier le temps \mathcal{K} où elle s'est produite.

L'implantation informatique des tests de Lombard (1987) et de notre nouveau test de détection dans la dépendance a été réalisée à l'aide du logiciel Matlab. Ces programmes se retrouvent à l'annexe A.

La méthodologie qui vient d'être décrite nous libère des hypothèses contraignantes imposées par les tests de Gombay & Horváth (1999) et Dias & Embrechts (2004). Il reste toutefois à établir la performance de notre méthode. Au chapitre suivant, la puissance du nouveau test est évaluée à l'aide d'une étude de simulations exhaustives.

CHAPITRE 5

ÉVALUATION DE LA PUISSANCE ET DE LA ROBUSTESSE DES NOUVEAUX TESTS

Ce chapitre a pour but d'évaluer la puissance de notre nouveau test pour détecter des ruptures dans la copule sous-jacente à une série d'observations bivariées. Deux situations seront considérées, à savoir

- (i) le cas de marges supposées constantes;
- (ii) le cas de marges sujettes à des ruptures.

Pour (i), il s'agit en quelque sorte d'étudier le test de Gombay & Horváth (1999) tel qu'expliqué à la section 4.3 dans sa formulation originale, mais pour la première fois sous l'angle des copules. Les résultats sont présentés et discutés à la section 5.1. Pour (ii), qui constitue la principale contribution de ce mémoire, les procédures de stabilisation des marges complètes sont appliquées. La puissance du test est évaluée à la section 5.2. La section 5.3, quant à elle, donnera une conclusion globale de cette étude de simulations.

5.1 Puissance de détection de points de rupture dans la dépendance sous des marges exemptes de ruptures

Cette section vise à estimer la puissance du test de Gombay & Horváth (1999), tel qu'élaboré à la section 4.3. Comme on l'a déjà fait remarquer, ce test permet la détection de points de rupture dans les copules, à condition de supposer que les lois marginales sont stables au cours du temps.

Dans les simulations effectuées, la dépendance entre deux variables aléatoires est caractérisée par le modèle, c'est-à-dire la forme de la copule, ainsi que par le degré de dépendance, mesuré par le tau de Kendall. Trois scénarios de rupture ont été considérés, à savoir un changement

- (1) dans la forme de la dépendance, mais avec le même tau de Kendall ;
- (2) du tau de Kendall pour une même famille de copules ;
- (3) simultané de forme de la dépendance et du tau de Kendall.

Tous ces scénarios sont exempts de rupture dans les lois marginales.

La puissance du test de détection a été évaluée sous trois modèles de copules, à savoir Clayton, Frank et Gumbel–Hougaard. Ces copules appartiennent à la famille des copules Archimédiennes. Se référer au chapitre 3 ainsi qu'à Genest & MacKay (1986) et Genest & Rivest (1993) pour de plus amples détails sur cette famille de copules. Les raisons pour lesquelles nous avons décidé

d'utiliser ces trois modèles de copules sont : (i) elles couvrent l'ensemble du spectre de la dépendance positive, c'est-à-dire qu'elles ont Π et M comme cas particuliers, (ii) elles sont populaires et (iii) elles permettent de modéliser des types de dépendance différents. Le tableau 5.1 résume d'ailleurs quelques caractéristiques de ces familles de copules.

Table 5.1: Description de trois copules Archimédiennes

Copule	$\phi_\theta(t)$	$\mathcal{C}_\theta(u, v)$	Θ
Clayton	$\frac{t^{-\theta} - 1}{\theta}$	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$[0, \infty)$
Frank	$\ln \left(\frac{e^{\theta x} - 1}{e^\theta - 1} \right)$	$-\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)$	$[0, \infty)$
Gumbel-Hougaard	$(-\ln t)^\theta$	$\exp(-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta})$	$[0, 1]$

La copule de Gumbel-Hougaard appartient aussi à la classe des copules de valeurs extrêmes, tel que remarqué par Capéraà et al. (2000). Cette famille est fréquemment utilisée en hydrologie et en finance. Une copule appartenant à cette classe est de la forme générale

$$\mathcal{C}_A(u, v) = \exp \left\{ \log uv A \left(\frac{\log u}{\log uv} \right) \right\}$$

en terme d'une fonction de dépendance $A : [0, 1] \rightarrow [0, 1]$ telle que $A(0) = A(1) = 1$ et $A(t) \geq \max(t, 1 - t)$. Dans le cas particulier où $A(t) = 1$, on retrouve la copule d'indépendance $\Pi(u, v) = uv$.

La copule de Gumbel–Hougaard admet uniquement de la dépendance positive, en particulier les cas où le lien entre les valeurs élevées de (X, Y) est fort. Les copules de Clayton et Frank sont dites complètes au sens où elles permettent de rendre compte de tous les types de dépendance.

5.1.1 Rupture dans le type de dépendance des copules pour un même degré de dépendance

Les résultats de simulation pour cette partie sont présentés au tableau 5.2. On remarque d'emblée que les puissances tendent à augmenter en fonction de la taille des échantillons, sauf pour une rupture qui fait passer la nature de la dépendance d'une copule de Clayton à une copule de Frank.

Le passage de la copule de Clayton à celle de Frank est faiblement détectée et ce, même lorsqu'on augmente la taille de l'échantillon. Les puissances de détection restent toutes inférieures à 15.40%, ce qui pourrait s'expliquer par la similitude de ces deux copules. Les puissances de détection pour les copules de Clayton à Gumbel–Hougaard et de Frank à Gumbel–Hougaard sont bonnes pour les cas où le degré de dépendance est faible ou élevé.

Pour l'ensemble des copules étudiées, les meilleures puissances sont obtenues lorsque le moment de rupture \mathcal{K} survient à 50% des observations. Cependant, pour une même paire de copules, les puissances ne sont pas symétriques par rapport au moment de la rupture. Par exemple, pour les copules de Clayton à Gumbel–Hougaard, les puissances pour une détection de rupture qui a lieu à 25% des données sont très éloignées de celles obtenues pour une rupture à 75%

des données. Cet aspect gagnerait à être étudié théoriquement. Finalement, c'est lorsque $\tau = 0.5$ que les puissances sont les plus faibles, peu importe le moment \mathcal{K} de la rupture. Les raisons de ce comportement du test mériteraient également d'être considérées dans un cadre formel.

Table 5.2: Puissance, estimée par 1000 répétitions, du test de détection de rupture de copules pour des changements dans le type de dépendance mais pour un tau de Kendall stable. Panneau supérieur : $n = 50$; panneau inférieur : $n = 100$.

C_1	C_2	\mathcal{K}	τ				
			.1	.25	.50	.75	.90
Clayton	Frank	25%	10.10	10.50	7.80	15.40	9.70
		50%	8.30	10.20	9.10	14.10	9.80
		75%	8.70	7.90	8.30	13.90	11.70
Clayton	Gumbel	25%	42.20	20.00	9.60	13.70	24.50
		50%	52.90	24.20	8.10	26.60	69.30
		75%	29.50	15.30	7.30	27.60	62.00
Frank	Gumbel	25%	42.20	21.20	8.80	14.70	27.10
		50%	52.40	24.50	10.30	27.50	72.00
		75%	26.40	15.60	8.30	24.20	62.90
Clayton	Frank	25%	5.90	6.30	5.90	7.20	10.70
		50%	9.00	6.10	6.20	7.60	11.40
		75%	7.00	9.90	6.10	7.30	8.80
Clayton	Gumbel	25%	64.60	28.80	7.30	20.20	69.30
		50%	83.10	42.40	6.00	52.50	99.20
		75%	52.30	22.80	5.50	40.10	95.20
Frank	Gumbel	25%	66.60	32.30	6.90	22.00	70.60
		50%	82.90	43.00	4.80	54.90	99.20
		75%	53.80	24.90	5.60	41.20	94.70

5.1.2 Rupture au niveau du degré de dépendance des copules pour un même type de dépendance

Les résultats de simulation pour cette partie sont présentés au tableau 5.3. De façon générale, les puissances sont fortement influencées par la variation du degré de dépendance. Plus cet écart est élevé, meilleurs sont les résultats. De plus, le passage à une dépendance plus forte ou plus faible a également un impact. Si le degré de dépendance augmente après la rupture, les puissances de détection seront meilleures si la rupture a lieu à 25% des observations plutôt qu'à 75%. Le contraire est aussi vrai : quand le degré de dépendance passe d'un niveau élevé à un niveau plus faible, les puissances sont plus élevées pour une rupture à 75% qu'à 25%. Par exemple, prenons la copule de Clayton ($n = 50$) qui subit un changement de tau de Kendall de 0.25 à 0.75. La puissance observée pour une rupture à $\mathcal{K} = 25\%$ est de 69.90%, ce qui est plus élevé que 44.50%, la puissance observée pour une rupture à 75% des données. Cette constatation confirme que la détection de points de rupture n'est pas symétrique par rapport au moment de cette rupture.

La détection des ruptures pour la copule de Gumbel–Hougaard ne s'améliore pas avec la taille de l'échantillon contrairement aux deux autres copules. En ce qui concerne les copules de Clayton et de Frank, les puissances de détection sont élevées, en particulier pour une variation des degrés de 0.25 à 0.75 et inversement. Encore une fois, les meilleures puissances de détection sont obtenues lorsque le moment de rupture est situé à 50% des données.

Table 5.3: Puissance, estimée par 1000 répétitions, du test de détection de rupture de copules pour des changements dans le tau de Kendall mais sous des modèles de copules stables. Panneau supérieur : $n = 50$; panneau inférieur : $n = 100$.

C	\mathcal{K}	(τ_1, τ_2)					
		$(.25, .5)$	$(.25, .75)$	$(.5, .75)$	$(.75, .5)$	$(.75, .25)$	$(.5, .25)$
Clayton	25%	22.50	69.90	29.80	17.00	50.70	20.20
	50%	28.30	82.40	35.70	37.20	82.70	26.50
	75%	17.10	44.50	17.50	30.90	65.00	20.70
Frank	25%	24.20	72.60	33.50	19.40	56.00	16.50
	50%	30.20	86.70	39.50	39.00	87.00	29.70
	75%	15.30	47.60	18.40	33.60	67.50	23.50
Gumbel	25%	8.30	11.50	7.50	8.00	9.10	8.30
	50%	8.10	9.30	8.20	8.00	7.90	10.40
	75%	9.20	8.30	10.00	10.50	8.70	8.70
Clayton	25%	30.90	95.00	51.30	35.20	88.90	26.90
	50%	49.00	99.20	67.00	66.50	98.60	46.70
	75%	28.00	89.00	34.50	46.10	94.60	33.70
Frank	25%	35.50	97.40	59.60	44.60	93.50	27.10
	50%	53.70	100.00	77.80	77.80	99.50	49.80
	75%	32.50	92.50	41.30	54.30	97.60	37.30
Gumbel	25%	6.40	11.40	7.60	6.50	6.70	6.30
	50%	6.20	8.50	8.00	7.70	9.10	7.20
	75%	5.80	5.30	5.60	6.50	10.40	7.60

5.1.3 Rupture simultanée dans le type de dépendance ainsi qu'au niveau du degré de dépendance des copules

Les résultats de cette partie sont présentés au tableau 5.4. Comme le plan combine, en quelques sortes, les scénarios des deux sous-sections précédentes,

plusieurs résultats s'expliquent par ce qui a déjà été remarqué.

Les copules de Clayton et Frank offrent de faibles puissances dans les simulations de la section 5.1.1, alors qu'elles présentent chacune des puissances élevées dans les simulations de la section 5.1.2. Dans cette simulation, les résultats sont bons lorsque la variation des degrés de dépendance est élevée, que ce soit vers un degré de dépendance plus fort ou plus faible. Cette constatation nous amène à croire que seule la rupture dans le degré de dépendance est détectée ici.

Comme mentionné précédemment dans la première simulation, les passages des copules de Frank et Clayton vers celle de Gumbel-Hougaard présentent de faibles résultats pour un tau de Kendall de 0.5, mais de très bons pour des degrés de dépendance faibles ou élevés. On remarque cet effet encore ici, puisque les meilleures puissances sont observées pour des variations de degrés élevées.

Encore une fois, les résultats ne sont pas symétriques par rapport au moment de la rupture. Pour un passage vers un degré plus élevé, les meilleures puissances de détection surviennent lorsque la rupture se produit au début des observations, et vice-versa. Par exemple, considérons le cas des copules de Clayton à Frank ($n = 50$), qui subissent un changement de tau de Kendall de 0.25 à 0.75. La puissance est de 73.20% pour une rupture à 25% des données comparativement à 46.70% pour une rupture à 75%. Il serait intéressant de se questionner et d'explorer les raisons de ce phénomène, étant donné qu'il est remarqué dans d'autres simulations, par exemple à la section 5.1.2.

Table 5.4: Puissance, estimée par 1000 répétitions, du test de détection de rupture de copules pour des changements dans le tau de Kendall et la copule. Panneau supérieur : $n = 50$; panneau inférieur : $n = 100$.

C_1	C_2	\mathcal{K}	(τ_1, τ_2)					
			$(.25, .5)$	$(.25, .75)$	$(.5, .75)$	$(.75, .5)$	$(.75, .25)$	$(.5, .25)$
Clayton	Frank	25%	23.20	73.20	31.00	18.50	54.30	18.30
		50%	27.70	84.40	36.40	39.80	86.20	29.00
		75%	17.60	46.70	16.80	28.10	69.90	20.10
Clayton	Gumbel	25%	22.10	28.30	9.50	16.00	22.10	9.10
		50%	26.80	37.20	7.90	39.60	40.50	6.90
		75%	18.00	21.00	8.50	31.30	34.30	8.60
Frank	Gumbel	25%	22.80	32.90	9.60	19.00	18.70	9.80
		50%	29.70	35.70	9.10	39.10	38.50	7.50
		75%	16.10	21.30	8.50	30.90	34.10	7.90
Clayton	Frank	25%	35.60	95.70	52.20	45.30	89.90	30.80
		50%	50.70	99.90	68.70	71.90	98.90	45.70
		75%	26.00	89.30	34.00	56.00	94.80	34.20
Clayton	Gumbel	25%	32.80	47.60	8.60	29.80	41.80	5.80
		50%	47.90	62.10	7.10	66.30	72.40	6.20
		75%	27.10	33.10	6.00	50.90	54.10	5.90
Frank	Gumbel	25%	33.30	43.00	8.40	37.00	35.40	5.90
		50%	52.50	65.10	7.50	72.00	77.40	5.50
		75%	30.40	35.20	4.90	53.00	62.90	7.40

5.2 Puissance de détection de points de rupture dans la dépendance sous des ruptures dans les marges

5.2.1 Description des scénarios considérés

Cette partie de l'étude a été réalisée uniquement pour les copules de Clayton et de Frank étant donné la lourdeur de la tâche. Dans la suite, ces modèles

sont identifiés respectivement par \mathcal{C}^{CL} et \mathcal{C}^{F} . Rappelons que le but de ce nouveau test est de détecter des points de rupture dans la copule tout en permettant aux lois marginales de se modifier au cours du temps. Trois scénarios, dénotés par A , B et C , ont donc été élaborés pour vérifier la puissance associée à ce nouveau test de détection. Tous ces scénarios incluent deux ruptures qui se produisent simultanément à 50% des observations dans le modèle de copules ainsi que dans le degré de dépendance, mesuré par $\tau \in \{.2, .4, .6\}$. Seuls des échantillons de taille $n = 50$ ont été considérés, compte tenu du long travail numérique à effectuer pour chaque scénario.

Le scénario *Référence* sert uniquement de base comparative à cette étude, puisqu'aucun changement dans les marges n'est introduit. Comparativement à la section précédente, cependant, la procédure complète est appliquée. Autrement dit, on applique les transformations de marges si la procédure détecte des changements à l'intérieur de celles-ci.

Le scénario A , quant à lui, étudie la puissance de détection de point de rupture dans la loi de dépendance où différentes ruptures de moyennes se produisent autant dans la loi individuelle de X que de Y . Les changements de moyennes dans la loi de X sont abrupts entre $\mathcal{K}_{\mu_1} = 25\%$ et $\mathcal{K}_{\mu_2} = 25\% + 1$, alors que les changements de moyennes dans la loi de Y sont graduels entre $\mathcal{K}_{\nu_1} = 25\%$ et $\mathcal{K}_{\nu_2} = 75\%$. Le scénario A se divise en quatre sous-scénarios comme suit :

$$A_1. \quad |\mu_1 - \mu_2| = 0.5, \quad |\nu_1 - \nu_2| = 0.5$$

$$A_2. \quad |\mu_1 - \mu_2| = 0.5, \quad |\nu_1 - \nu_2| = 1.5$$

$$A_3. |\mu_1 - \mu_2| = 1.5, |\nu_1 - \nu_2| = 0.5$$

$$A_4. |\mu_1 - \mu_2| = 1.5, |\nu_1 - \nu_2| = 1.5$$

Le scénario *B* étudie la puissance de détection lorsque des changements dans la variance se produisent. Encore une fois, les changements de variances dans la loi de X sont abrupts entre $\mathcal{K}_{\sigma_1} = 25\%$ et $\mathcal{K}_{\sigma_2} = 25\% + 1$, alors que les changements de variances dans la loi de Y sont graduels entre $\mathcal{K}_{\gamma_1} = 25\%$ et $\mathcal{K}_{\gamma_2} = 75\%$. Le scénario *B* se divise en quatre sous-scénarios comme suit :

$$B_1. \sigma_1/\sigma_2 = 1.5, \gamma_1/\gamma_2 = 1.5$$

$$B_2. \sigma_1/\sigma_2 = 1.5, \gamma_1/\gamma_2 = 2.5$$

$$B_3. \sigma_1/\sigma_2 = 2.5, \gamma_1/\gamma_2 = 1.5$$

$$B_4. \sigma_1/\sigma_2 = 2.5, \gamma_1/\gamma_2 = 2.5$$

Le scénario *C* étudie la puissance de détection du nouveau test lorsque les lois X et Y subissent des changements dans leurs moyennes et leurs variances. Cette fois-ci, les changements de moyennes dans la loi de X et la loi de Y sont abrupts entre $\mathcal{K}_{\mu_1} = 25\%$ et $\mathcal{K}_{\mu_2} = 25\% + 1$ et entre $\mathcal{K}_{\nu_1} = 75\%$ et $\mathcal{K}_{\nu_2} = 75\% + 1$ respectivement. Les changements de variances dans la loi de X et de Y sont abrupts entre $\mathcal{K}_{\sigma_1} = 75\%$ et $\mathcal{K}_{\sigma_2} = 75\% + 1$ et entre $\mathcal{K}_{\gamma_1} = 25\%$ et $\mathcal{K}_{\gamma_2} = 25\% + 1$, respectivement. Le scénario *C* se divise ainsi :

$$C_1. |\mu_1 - \mu_2| = 0.5, |\nu_1 - \nu_2| = 0.5, \sigma_1/\sigma_2 = 1.5, \gamma_1/\gamma_2 = 1$$

$$C_2. |\mu_1 - \mu_2| = 0.5, |\nu_1 - \nu_2| = 0.5, \sigma_1/\sigma_2 = 2.5, \gamma_1/\gamma_2 = 1$$

$$C_3. |\mu_1 - \mu_2| = 0.5, |\nu_1 - \nu_2| = 1.5, \sigma_1/\sigma_2 = 1.5, \gamma_1/\gamma_2 = 1$$

$$C_4. |\mu_1 - \mu_2| = 0.5, |\nu_1 - \nu_2| = 1.5, \sigma_1/\sigma_2 = 2.5, \gamma_1/\gamma_2 = 1$$

$$C_5. |\mu_1 - \mu_2| = 1.5, |\nu_1 - \nu_2| = 0.5, \sigma_1/\sigma_2 = 1.5, \gamma_1/\gamma_2 = 1$$

$$C_6. |\mu_1 - \mu_2| = 1.5, |\nu_1 - \nu_2| = 0.5, \sigma_1/\sigma_2 = 2.5, \gamma_1/\gamma_2 = 1$$

$$C_7. |\mu_1 - \mu_2| = 1.5, |\nu_1 - \nu_2| = 1.5, \sigma_1/\sigma_2 = 1.5, \gamma_1/\gamma_2 = 1$$

$$C_8. |\mu_1 - \mu_2| = 1.5, |\nu_1 - \nu_2| = 1.5, \sigma_1/\sigma_2 = 2.5, \gamma_1/\gamma_2 = 1$$

$$C_9. |\mu_1 - \mu_2| = 1.5, |\nu_1 - \nu_2| = 1.5, \sigma_1/\sigma_2 = 2.5, \gamma_1/\gamma_2 = 2.5$$

Il est à noter que μ_x^D signifie l'ampleur de la rupture de la moyenne pour X , et que μ_y^D est l'ampleur de la rupture de la moyenne pour Y . De même, σ_x^D signifie le ratio entre les écart-types avant et après une rupture pour X , et σ_y^D est le ratio entre les écart-types avant et après une rupture pour Y . Les valeurs utilisées pour les différences de moyennes sont basées sur les résultats des tableaux 2.3 et 2.4 de telle sorte qu'on considère deux cas de rupture : une faible différence et une grande différence de moyennes. De plus, les valeurs utilisées pour les différences de variances sont basées sur les résultats des tableaux 2.5 et 2.6 de telle sorte qu'on considère aussi deux cas de rupture : une faible différence et une grande différence de variances.

5.2.2 Résultats et interprétations

Les tableaux 5.5 et 5.6 présentent les résultats pour les scénarios A , B et C décrits à la sous-section précédente. D'abord, les résultats obtenus pour

le scénario *Référence* nous indique que la procédure éprouve de la difficulté à conserver son seuil de 5% si aucune rupture de dépendance n'a lieu. En effet, les puissances varient entre 6.90 et 10.10. Ces puissances peuvent être expliquées probablement par la petite taille d'échantillon étudiée. Le scénario A offre, similairement au scénario *Référence*, des puissances situées entre 6.40 et 11.70 lorsqu'il n'y a aucun changement au niveau de la copule. Les meilleures puissances de détection, pour le scénario A, sont observées lorsque la rupture dans le degré de dépendance passe de 0.2 à 0.6. Il en est de même, en général, pour tous les autres scénarios à l'étude.

Pour le scénario B, lorsqu'il n'y a aucun changement de dépendance au niveau de la copule, on observe que le sous-scénario B_1 conserve généralement mieux son seuil que les autres sous-scénarios. À noter que le ratio des variances pour B_1 est le plus faible par rapport aux autres sous-scénarios. On remarque que les puissances entre les scénarios A et B sont similaires, ce qui indique que la procédure détecte et stabilise aussi bien les ruptures de moyennes que de variances pour ensuite détecter les ruptures dans la dépendance.

En ce qui concerne le scénario C, on remarque que les puissances lorsque le degré de dépendance passe de 0.2 à 0.6 sont légèrement plus faibles que pour les scénarios A et B. Ceci indique, probablement, que la double détection et stabilisation des moyennes et des variances a une certaine influence sur la détection des ruptures dans la dépendance. À partir du sous-scénario C_6 , les puissances tendent à augmenter même lorsqu'il n'y a aucun changement dans la copule. À noter que les variations de moyennes et de variances pour chacune des lois marginales sont plus élevées à partir de ce sous-scénario

aussi. Le sous-scénario C_9 présente des seuils assez élevés pour les cas où il n'y a aucune rupture au niveau de la copule. Ce sous-scénario est celui, aussi, qui présente la plus grande différence de moyennes et de variances dans les lois individuelles de X et de Y . Cette difficulté à conserver son seuil s'explique probablement par une taille d'échantillon trop petite.

Table 5.5: Puissance de la procédure complète pour la détection de points de rupture de copules sous le scénario de référence ainsi que sous les scénarios A et B pour des tailles d'échantillons de $n = 50$

C_1	C_2	Référence	Scenario A : changes in the means				Scenario B : changes in the variances			
			A_1	A_2	A_3	A_4	B_1	B_2	B_3	B_4
$C_{.2}^{CL}$	$C_{.2}^{CL}$	8.30	7.80	9.20	8.70	8.00	8.80	10.00	10.20	10.20
$C_{.2}^{CL}$	$C_{.4}^{CL}$	16.10	17.70	14.70	21.90	19.00	16.30	17.40	18.40	22.40
$C_{.2}^{CL}$	$C_{.6}^{CL}$	39.50	45.10	36.90	43.50	39.90	42.20	40.90	42.00	43.20
$C_{.4}^{CL}$	$C_{.4}^{CL}$	8.40	8.50	8.30	9.60	7.70	8.50	10.60	10.60	10.20
$C_{.4}^{CL}$	$C_{.6}^{CL}$	16.20	20.70	20.60	22.40	19.60	17.60	21.90	21.90	23.20
$C_{.6}^{CL}$	$C_{.6}^{CL}$	9.80	10.10	11.70	8.40	10.20	9.90	12.50	10.80	13.00
$C_{.2}^{CL}$	$C_{.2}^F$	9.40	11.80	8.00	7.50	9.60	13.30	8.60	10.20	10.80
$C_{.2}^{CL}$	$C_{.4}^F$	19.30	18.70	16.20	19.90	18.20	19.20	18.60	20.50	21.60
$C_{.2}^{CL}$	$C_{.6}^F$	48.30	46.70	36.90	46.60	40.10	42.30	45.10	45.00	43.60
$C_{.4}^{CL}$	$C_{.4}^F$	8.50	7.70	10.20	10.10	10.30	8.30	11.90	10.70	12.20
$C_{.4}^{CL}$	$C_{.6}^F$	17.80	19.90	20.80	24.80	20.10	17.90	24.00	22.00	23.30
$C_{.6}^{CL}$	$C_{.6}^F$	10.10	9.50	10.00	9.70	9.00	9.90	11.30	13.60	11.70
$C_{.2}^F$	$C_{.2}^F$	8.30	7.70	7.80	7.00	6.40	9.30	9.00	9.40	9.60
$C_{.2}^F$	$C_{.4}^F$	16.60	16.90	15.10	20.50	17.10	17.60	21.70	19.00	22.20
$C_{.2}^F$	$C_{.6}^F$	43.90	48.70	38.00	45.40	39.00	45.80	45.10	48.0	48.50
$C_{.4}^F$	$C_{.4}^F$	6.90	8.30	9.00	8.80	8.30	8.20	8.20	9.40	9.80
$C_{.4}^F$	$C_{.6}^F$	18.80	22.00	19.00	21.50	18.90	20.00	23.90	23.20	24.20
$C_{.6}^F$	$C_{.6}^F$	9.20	7.00	9.40	9.30	10.70	9.50	10.70	12.00	13.10

Table 5.6: Puissance de la procédure complète pour la détection de points de rupture de copules sous le scénario C pour des tailles d'échantillons de $n = 50$

C_1	C_2	Scenario D : changes in the means and variances								
		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
C_2^{CL}	C_2^{CL}	8.30	9.00	7.90	10.80	9.00	11.60	10.50	12.50	29.40
C_2^{CL}	C_4^{CL}	16.80	17.00	16.90	18.80	18.40	21.60	16.90	23.00	26.20
C_2^{CL}	C_6^{CL}	33.10	37.90	34.50	42.00	35.80	40.10	37.70	46.50	34.90
C_4^{CL}	C_4^{CL}	9.50	8.90	9.50	10.30	12.00	13.00	10.60	11.50	16.80
C_4^{CL}	C_6^{CL}	15.60	18.40	15.60	16.20	18.00	25.80	17.60	25.60	20.30
C_6^{CL}	C_6^{CL}	9.50	9.80	9.30	8.40	10.30	15.60	9.00	14.00	11.20
C_2^{CL}	C_2^F	8.20	8.70	8.30	9.90	8.20	10.10	9.00	11.10	27.00
C_2^{CL}	C_4^F	14.90	17.40	14.70	17.30	17.70	18.80	15.50	22.20	27.80
C_2^{CL}	C_6^F	33.30	35.50	33.30	36.40	36.00	40.50	36.50	42.10	33.20
C_4^{CL}	C_4^F	8.40	8.90	7.90	9.30	10.60	12.50	10.50	12.40	17.40
C_4^{CL}	C_6^F	15.10	15.30	12.70	18.40	17.70	22.90	16.70	22.80	21.30
C_6^{CL}	C_6^F	8.60	8.50	6.40	7.80	10.80	16.00	8.90	14.50	10.90
C_2^F	C_2^F	9.60	8.20	8.90	9.40	9.60	11.40	10.20	10.90	26.20
C_2^F	C_4^F	14.20	16.10	12.80	17.90	16.90	18.60	15.60	23.50	28.80
C_2^F	C_6^F	33.20	36.50	34.30	36.80	34.90	38.00	37.40	42.20	34.40
C_4^F	C_4^F	9.00	9.30	7.90	7.90	8.70	12.20	9.60	12.60	18.50
C_4^F	C_6^F	14.10	14.80	12.20	17.70	16.50	21.10	15.50	25.30	19.30
C_6^F	C_6^F	8.80	8.60	6.60	6.90	9.00	15.40	9.10	14.00	13.10

5.3 Conclusion générale sur l'étude de puissance

La première partie de ce chapitre a été consacrée à l'évaluation de notre méthodologie lorsque les marges sont supposées constantes. Les ruptures qui se produisent à 50% des observations sont les mieux détectées pour l'ensemble des scénarios. En général, les puissances tendent à augmenter en fonction de la taille d'échantillon. Par contre, deux cas ne respectent pas cette tendance, en particulier, le passage de la copule de Clayton à celle de Frank pour un même tau de Kendall (section 5.1.1) et la copule de Gumbel-Hougaard pour une rupture de tau de Kendall (section 5.1.2). Tel qu'attendu, plus l'écart du tau de Kendall est élevé, meilleures sont les puissances, peu importe le lieu \mathcal{K} de la rupture.

La deuxième partie de ce chapitre s'est attardée à la puissance en présence de ruptures dans les marges. En général, la procédure éprouve de la difficulté à conserver son seuil. La méthodologie offre néanmoins les meilleures puissances de détection lorsque la rupture du tau de Kendall passe de 0.2 à 0.6. De plus, à l'aide du scénario C, nous avons pu remarquer que le ratio des variations affecte la détection de rupture dans la dépendance. Finalement, la double détection en moyenne et en variance a aussi un impact sur les puissances finales.

Au chapitre suivant, la nouvelle méthodologie est utilisée pour détecter de possibles ruptures de dépendance en lien avec l'écoulement annuel moyen de quelques bassins du Québec.

CHAPITRE 6

APPLICATION SUR DES DONNÉES HYDROLOGIQUES RÉELLES

6.1 Présentation des données

Dans ce chapitre, la procédure complète de détection de rupture sera appliquée à des observations simulées issues du *Modèle régional canadien du climat*. Voir, à ce sujet, Caya & Laprise (1999) et Music & Caya (2007). Ces données de simulation représentent les écoulements moyens annuels de différents bassins versants du Québec. La recherche de l'existence de points de rupture pourrait nous renseigner sur les causes susceptibles d'induire ces changements.

Plus spécifiquement, l'application utilisera les données simulées *acu* concernant l'écoulement annuel moyen du Modèle régional canadien du climat MRCC 4.1.1. Les simulations sont pilotées aux frontières par les champs atmosphériques de deux simulations du modèle canadien couplé de climat global, à savoir les quatrième et cinquième membres de MCCG3. Voir Flato

& Boer (2001) et Scinocca et al. (2008) pour plus de détails. Les simulations globale et régionale ont été produites en suivant les scénarios projetés de gaz à effet de serre et d'aérosols du GIEC SRES A2. Le domaine de pilotage est centré sur le Québec avec 112×88 points de grille et une résolution horizontale de quarante-cinq kilomètres. Les données du MRCC4.1.1 ont été générées et fournies par le consortium *Ouranos*.

Toute la période de simulation disponible, à savoir les cent quarante-deux années allant de 1958 à 2099, sera utilisée. Normalement, les années 1958, 1959 et 1960 auraient dû être enlevées de l'étude car elles représentent la période d'ajustement du MRCC4.1.1. Cependant, la présence de ces années n'affectent pas de façon significative les résultats obtenus.

Les simulations de l'écoulement moyen annuel sont exprimées en mm/jour. La carte qui localise les différents bassins du Québec est présentée à la figure 6.1. Voici la liste complète des bassins.

- | | |
|----------------------------------|-------------------------------|
| 1. ARN Rivière Arnaud | 12. MEL Rivière aux Mélèzes |
| 2. BAL Rivière À la Baleine | 13. MOI Rivière Moisie |
| 3. BEL Rivière Bell | 14. NAT Rivière Natashquan |
| 4. BOM Bersimis-Outardes-Manic | 15. PYR Rivière Caniapiscau |
| 5. CAN Réservoir Caniapiscau | 16. RDO Rivière des Outaouais |
| 6. CHU Réservoir Churchill Falls | 17. ROM Rivière Romaine |
| 7. FEU Rivière aux Feuilles | 18. RUP Rivière Rupert |
| 8. GEO Rivière George | 19. SAG Lac Saint-Jean |
| 9. GRB Grande riv. de la Baleine | 20. STM Rivière Saint-Maurice |
| 10. LGR La Grande Rivière | 21. WAS Rivière Waswanipi |
| 11. MAN Réservoir Manic 5 | |

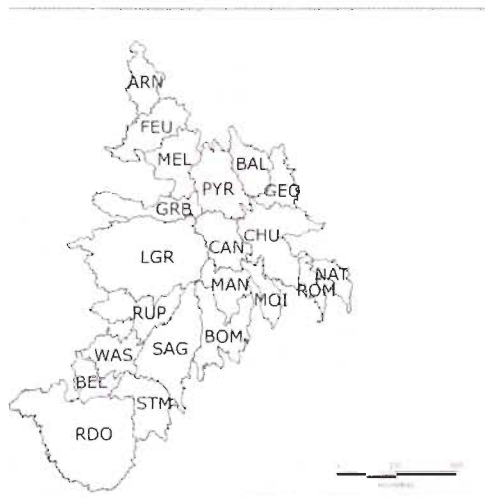


Figure 6.1: Bassins versants du Québec

6.2 Détection de ruptures pour les bassins *À la Baleine* et *Georges*

L'illustration suivante porte sur les bassins BAL et GEO, c'est-à-dire les rivières *À la Baleine* et *George*. Le choix de ces deux bassins versants est motivé par leur proximité géographique et donc par la dépendance probable qui les unit. Les variables qui seront étudiées sont les écoulements annuels moyens de ces deux rivières.

6.2.1 Analyses descriptives des échantillons

La figure 6.2 présente les écoulements annuels moyens sur la rivière *À la Baleine* pour l’horizon 1958 à 2099. On peut détecter, visuellement, une tendance à la hausse à partir de l’année 2050. La table 6.1 présente quelques statistiques descriptives, à savoir la moyenne, l’écart-type, le minimum, le maximum et la médiane des données. On retrouve également les coefficients d’asymétrie et d’aplatissement. Ceux-ci indiquent que la distribution est légèrement asymétrique et plus aplatie que la loi normale. Ensuite, on a voulu vérifier si ces simulations sont distribuées selon la loi normale. La valeur de la statistique du test de Lilliefors (1967), conçu à cette fin, ne permet pas de réfuter cette hypothèse. Ceci est confirmé par la figure 6.3 qui représente la droite de Henry. Sur celle-ci, les simulations sont tracées en fonction des probabilités empiriques de l’échantillon. Pour un échantillon normal, les données simulées devraient être alignées. On voit que c’est bien le cas ici. Pour plus de détails sur l’utilisation de la droite de Henry en hydrologie, voir par exemple Haché et al. (1999).

Table 6.1: Statistiques descriptives des écoulements annuels moyens sur la rivière *À la Baleine* entre les années 1958 et 2099

Écoulement minimal	0.7298	Écoulement médian	1.2632
Écoulement maximal	1.9235	Coefficient de variation	0.1804
Écoulement moyen	1.2702	Coefficient d’asymétrie	0.2451
Écart-type	0.2291	Coefficient d’aplatissement	2.6461

La figure 6.4 présente les écoulements annuels moyens sur la rivière *George*.

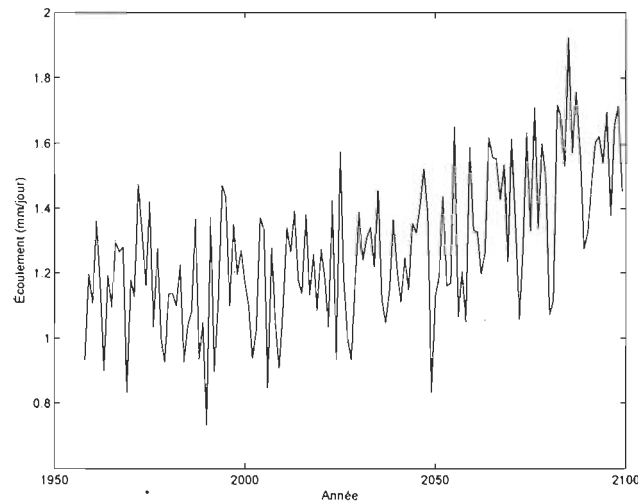


Figure 6.2: Écoulements annuels moyens sur la rivière *À la Baleine* entre les années 1958 et 2099

À l'instar de la rivière *À la Baleine*, on perçoit une légère dérive à la hausse vers l'an 2050. La table 6.2 présente quelques statistiques descriptives. On remarque à nouveau une légère asymétrie. Le coefficient d'aplatissement est légèrement plus élevé que celui de la loi normale. À noter, cependant, que l'hypothèse de normalité est acceptée par le test de Lilliefors (1967) et confirmée par la droite de Henry présentée à la figure 6.5.

La figure 6.6 présente le nuage de points des rangs normalisés associés aux écoulements des rivières *À la Baleine* et *George*. Il semble exister un lien assez fort entre les deux variables. Ceci est confirmé par le calcul du rho de Spearman et du tau de Kendall. En effet, on a $r_S = 0.8468$ et $\tau_n = 0.6614$, ces deux valeurs étant significativement supérieures à 0 au seuil $\alpha = 0.05$.

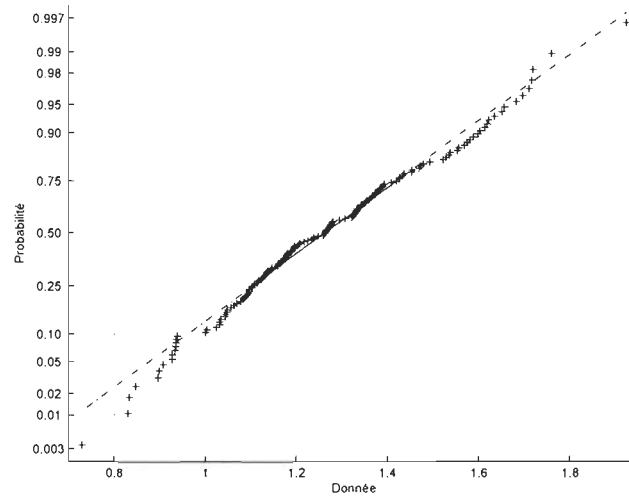


Figure 6.3: Droite de Henry des écoulements annuels moyens sur la rivière *À la Baleine* entre les années 1958 et 2009

Table 6.2: Statistiques descriptives des écoulements annuels moyens sur la rivière *George* entre les années 1958 et 2009

Écoulement minimal	0.7348	Écoulement médian	1.3912
Écoulement maximal	2.0682	Coefficient de variation	0.1682
Écoulement moyen	1.3919	Coefficient d'asymétrie	0.1510
Écart-type	0.2341	Coefficient d'aplatissement	3.1293

6.2.2 Détection de points de rupture dans les marges

L'application de la procédure complète de détection de rupture dans la structure de dépendance consiste d'abord à vérifier la stabilité des lois marginales. Le test de Lombard (1987) basé sur la statistique S_{n1} pour la détection d'un

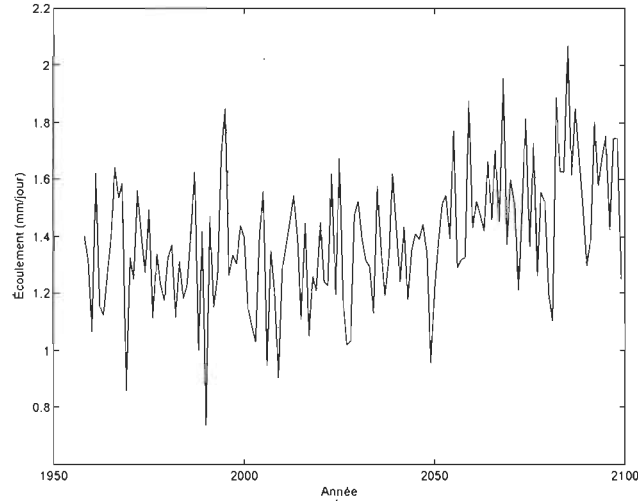


Figure 6.4: Écoulements annuels moyens sur la rivière *George* entre les années 1958 et 2099

point de rupture dans la moyenne et la variance sera donc effectué pour les deux bassins. Le choix de S_{n1} s'appuie sur les bons résultats obtenus suite à l'étude des différents scénarios élaborés au chapitre 2.

La valeur observée de la statistique de test pour la détection de point de rupture dans la moyenne pour la rivière *À la Baleine* est de 0.4143, ce qui est supérieur à la valeur critique de rejet de \mathcal{H}_0 , à savoir 0.0402. L'hypothèse nulle est alors rejetée et ce bassin présente donc une rupture significative au niveau de sa moyenne. Selon l'estimateur de Lombard (1987), les points de rupture entre lesquels la moyenne se modifie sont les années 2021 et 2094. Ce grand intervalle indique que la rupture est progressive. Une stabilisation de la moyenne est donc effectuée en utilisant $\hat{\mu}_1 = 1.1589$ et $\hat{\mu}_2 = 1.5732$. À noter qu'avant 2021, la moyenne était stable à 1.1589, mais qu'à partir

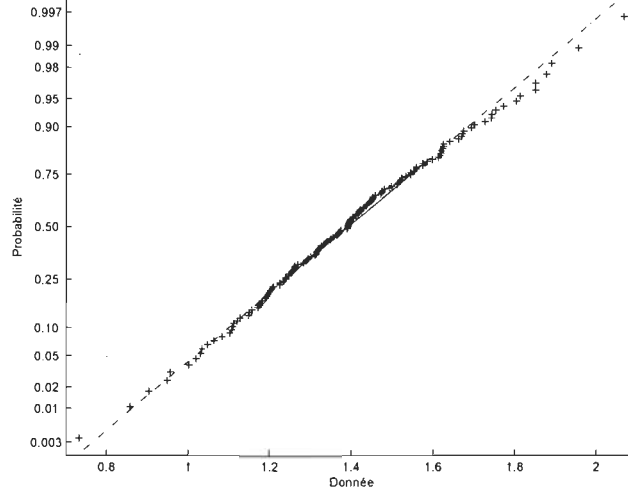


Figure 6.5: Droite de Henry des écoulements annuels moyens sur la rivière *George* entre les années 1958 et 2009

de 2021 elle a commencé à se modifier tranquillement jusqu'à l'année 2094 pour se stabiliser enfin à 1.5732. La figure 6.7 présente la série chronologique de ce bassin après la stabilisation de la moyenne. On note que la tendance croissante, remarquée à la figure 6.2, est éliminée.

Suite à la stabilisation de la moyenne, la statistique S_{n1} est calculée avec le score de Mood, c'est-à-dire $\phi_2(u) = (2u - 1)^2$, afin d'identifier d'éventuels points de rupture dans la variance. La valeur observée de la statistique est 0.0043, ce qui est en deçà de la valeur critique, soit 0.0403. Ainsi, la variance est constante au cours du temps et ceci assure que le comportement en moyenne et en variance de la rivière *À la Baleine* est homogène.

Le même procédé est appliqué au bassin de la rivière *George*. La statistique

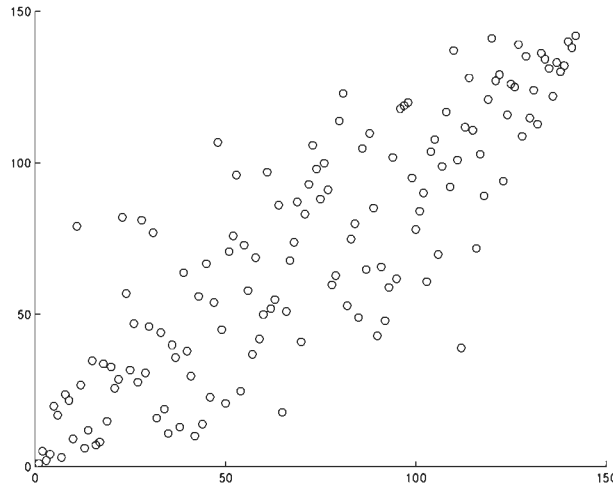


Figure 6.6: Nuage de points des rangs normalisés des écoulements annuels moyens des rivières *À la Baleine* et *George*

de test obtenue pour la détection de point de rupture dans la moyenne est de 0.2397, ce qui amène à rejeter \mathcal{H}_0 . Il y a donc un changement significatif de moyenne et les points de rupture sont estimés à 2050 et 2051. C'est donc dire que le changement est abrupt. La stabilisation de la moyenne est donc effectuée avec $\hat{\nu}_1 = 1.3111$ et $\hat{\nu}_2 = 1.5451$. La figure 6.8 représente la série chronologique une fois stabilisée. On voit que la tendance croissante, remarquée à la figure 6.4, a été éliminée.

À l'instar de la rivière *À la Baleine*, le test de détection de rupture dans la variance ne permet pas de conclure à un changement significatif. La table 6.3 résume les différents résultats obtenus de cette analyse.

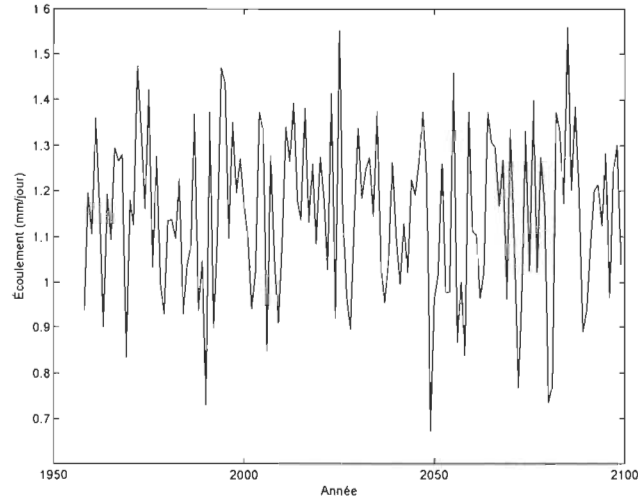


Figure 6.7: Écoulements annuels moyens sur la rivière *À la Baleine* entre les années 1958 et 2099 après stabilisation de la moyenne

6.2.3 Détection de ruptures dans la forme de la dépendance

Le test de détection de point de rupture dans la loi de dépendance, élaboré par Gombay & Horváth (1999) peut maintenant être appliqué aux séries dont les moyennes et les variances ont été stabilisées. On rappelle que les hypothèses à confronter sont $\mathcal{H}_0 : \mathcal{C}_1 = \mathcal{C}_2 = \dots = \mathcal{C}_n$ vs $\mathcal{H}_1 : \mathcal{C}_1 = \dots = \mathcal{C}_K \neq \mathcal{C}_{K+1} = \dots = \mathcal{C}_n$ pour un certain $K \in \{2, \dots, n-1\}$. L'hypothèse nulle est rejetée si la valeur observée de la statistique du test est supérieure à la valeur critique. Cette dernière est obtenue par une méthode de ré-échantillonnage expliquée au chapitre 4.

La valeur de la statistique du test de détection de rupture dans la dépendance entre les écoulements des deux bassins à l'étude est $\hat{M}_n = 0.2730$, alors que

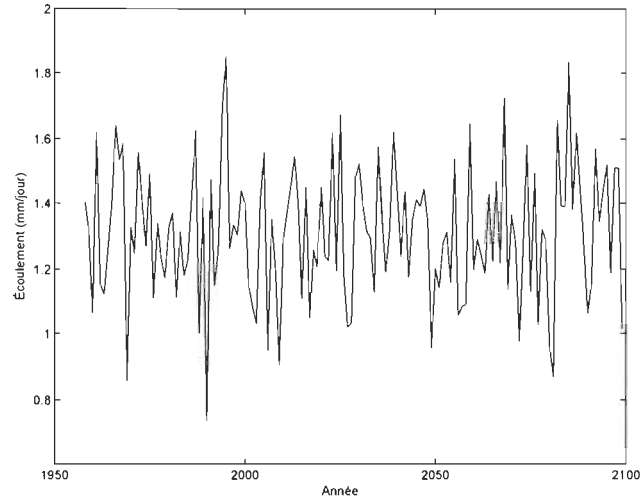


Figure 6.8: Écoulements annuels moyens sur la rivière *George* entre les années 1958 et 2099 après stabilisation de la moyenne

la valeur critique de rejet est estimée à 0.5464. L'hypothèse nulle de stabilité dans la fonction de dépendance est donc acceptée.

Il est également possible de visualiser les résultats de ce test par un outil graphique intéressant. En fait, il s'agit de rapporter, en fonction de $\mathcal{K} \in \{2, \dots, n-1\}$, toutes les valeurs des différences entre les taus de Kendall. Plus précisément, on considère

$$\frac{\mathcal{K}(n-\mathcal{K})}{n^{3/2}} \left| \hat{T}_{\mathcal{K}} - \hat{T}_{\mathcal{K}}^* \right|, \quad \mathcal{K} \in \{2, \dots, n-1\}.$$

La valeur maximale sur le graphique est simplement la statistique de test \hat{M}_n . La figure 6.9 en donne une illustration pour les simulations BAL et GEO, où la valeur critique est indiquée en rouge. On voit que l'hypothèse nulle fut loin d'être acceptée.

Table 6.3: Tests de détection de rupture en moyenne et en variance pour les bassins des rivières *À la Baleine* et *George*

Statistique		<i>À la Baleine</i>	<i>George</i>
Tests de ruptures de moyenne	Sn_1	0.4143	0.2397
	Valeur critique	0.0403	0.0403
	$(\hat{\mathcal{K}}_{\mu_1}, \hat{\mathcal{K}}_{\mu_2})$	(2021, 2094)	(2050, 2051)
	$\hat{\mu}_1$	1.1589	1.3111
	$\hat{\mu}_2$	1.5732	1.5451
Tests de rupture de variance	Sn_1	0.0043	0.0025
	Valeur critique	0.0403	0.0403
	$(\hat{\mathcal{K}}_{\sigma_1}, \hat{\mathcal{K}}_{\sigma_2})$	-	-
	$\hat{\sigma}_1$	-	-
	$\hat{\sigma}_2$	-	-

6.3 Analyse des ruptures dans la dépendance pour quelques bassins versants du Québec

Cette section présente les résultats obtenus suite à l'analyse des ruptures dans la loi de dépendance entre quelques paires de bassins du Québec. Les résultats de l'analyse des ruptures univariées pour chaque bassin sont d'abord

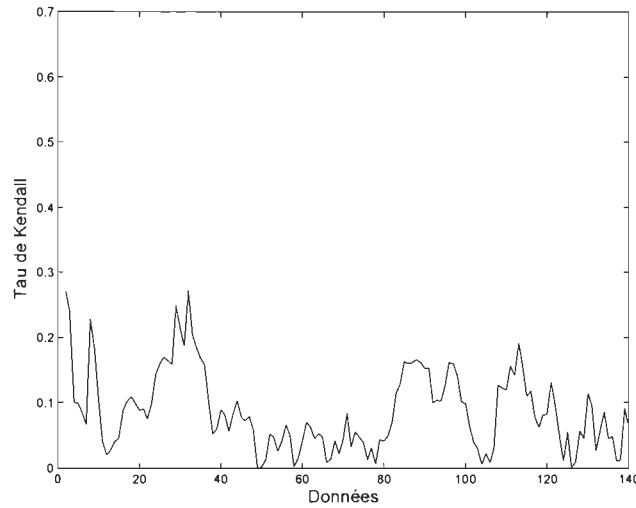


Figure 6.9: Outil graphique pour détecter une rupture de copule, illustration pour les bassins BAL et GEO

présentés. Ensuite, quelques paires de bassins sont sélectionnées pour l'analyse complète des ruptures dans leur loi de dépendance.

6.3.1 Détection des ruptures dans les marges

Pour vérifier l'existence de points de rupture dans la loi de dépendance, il faut d'abord identifier les points de rupture dans la moyenne et dans la variance. La table 6.4 présente donc les grandes lignes de l'analyse univariée de chaque bassin versant. On remarque que seuls les bassins des rivières *Moisie* (MOI), *Natashquan* (NAT) et *Romaine* (ROM) ne présentent pas de rupture de moyenne, ce qui signifie qu'aucune stabilisation ne doit être effectuée. Outre ces trois bassins, tous les autres ont subi une stabilisation de

moyenne. Ensuite, le test de détection de ruptures dans la variance indique que celle-ci est constante pour tous les bassins. Aucune transformation pour homogénéiser les variances des séries ne fut donc nécessaire.

Table 6.4: Statistiques Sn_1 et années de rupture en moyenne ($\hat{\mathcal{K}}_{\hat{\mu}_1}, \hat{\mathcal{K}}_{\hat{\mu}_2}$) et en variance ($\hat{\mathcal{K}}_{\hat{\sigma}_1}, \hat{\mathcal{K}}_{\hat{\sigma}_2}$) des vingt et un bassins versants du Québec

Bassins versants	Rupture en moyenne			Rupture en variance après stabilisation		
	Sn_1	$\hat{\mathcal{K}}_{\hat{\mu}_1}$	$\hat{\mathcal{K}}_{\hat{\mu}_2}$	Sn_1	$\hat{\mathcal{K}}_{\hat{\sigma}_1}$	$\hat{\mathcal{K}}_{\hat{\sigma}_2}$
ARN	0.5155	2006	2094	0.0331	—	—
BAL	0.4143	2021	2094	0.0043	—	—
BEL	0.1172	2034	2035	0.0142	—	—
BOM	0.0884	2064	2065	0.0125	—	—
CAN	0.3473	2027	2069	0.0274	—	—
CHU	0.1188	2056	2058	0.0020	—	—
FEU	0.6402	1987	2097	0.0198	—	—
GEO	0.2397	2050	2051	0.0025	—	—
GRB	0.7583	1982	2091	0.0399	—	—
LGR	0.4762	1989	2095	0.0290	—	—
MAN	0.1664	2064	2065	0.0003	—	—
MEL	0.6897	1983	2096	0.0270	—	—
MOI	0.0304	—	—	0.0271	—	—
NAT	0.0339	—	—	0.0037	—	—
PYR	0.5282	1988	2096	0.0086	—	—
RDO	0.0818	2055	2056	0.0114	—	—
ROM	0.0178	—	—	0.0056	—	—
RUP	0.3178	2022	2095	0.0132	—	—
SAG	0.1938	2032	2033	0.0018	—	—
STM	0.1054	2051	2052	0.0049	—	—
WAS	0.1475	2034	2035	0.0143	—	—

6.3.2 Analyse des points de rupture dans la dépendance

À présent que les marges sont stabilisées, on peut procéder à la détection de points de rupture dans la dépendance à l'aide du test de Gombay & Horváth (1999). Comme il y a $21 \times 20/2 = 210$ paires de bassins possibles, les analyses subséquentes s'attardent aux bassins du centre-est et du centre-ouest du Québec. La table 6.5 présente les valeurs de la statistique de test de détection de rupture dans la dépendance (\hat{M}_n) pour toutes les paires de bassins situées au centre-ouest du Québec. Cette région comprend les bassins *La Grande Rivière* (LGR), *Caniapiscou* (CAN), *Grande rivière à la Baleine* (GRB), *Manic5* (MAN) et *Rupert* (RUP). On constate qu'aucune rupture de dépendance n'est détectée pour les couples étudiés. La figure 6.10 permet de le constater visuellement pour le réservoir *Caniapiscou* (CAN) et le réservoir *Manic5* (MAN) à l'aide de l'outil graphique présenté à la section précédente.

Table 6.5: Statistiques de test et valeurs critiques estimées (triangle supérieur) et années de rupture de la dépendance, s'il y a lieu (triangle inférieur) pour les bassins versants du centre-ouest du Québec

Bassins	LGR	CAN	GRB	MAN	RUP
LGR		0.3576 (0.5044)	0.2799 (0.5683)	0.3406 (0.5977)	0.3256 (0.5295)
CAN			0.5067 (0.7133)	0.3784 (0.5925)	0.6070 (0.6884)
GRB				0.4595 (0.8285)	0.3314 (0.7831)
MAN					0.3737 (0.8032)
RUP					

La table 6.6 présente des résultats similaires à la table 6.5 pour les bassins du centre-est du Québec, à savoir *Churchill Falls* (CHU), *Moisie* (MOI), *Ro-*

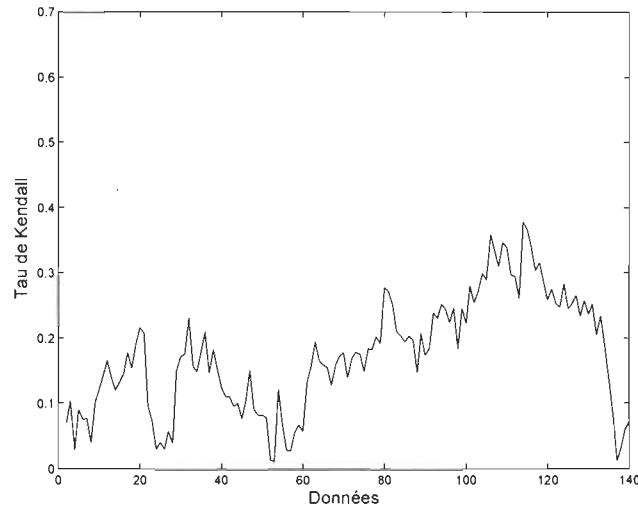


Figure 6.10: Détection graphique de la dépendance pour le réservoir *Canapiscou* et le réservoir *Manic5*

maine (ROM) et *Natashquan* (NAT). Ici, la copule entre certaines paires de bassins présente une rupture. C'est le cas, notamment, du couple *Churchill Falls* (CHU) et *Romaine* (ROM), où une rupture est identifiée en 2035. De même, la copule des rivières *Moisie* (MOI) et *Romaine* (ROM) se modifie en 2027. On peut d'ailleurs le constater en examinant la figure 6.11. Finalement, on remarque que le réservoir *Churchill Falls* (CHU) présente une rupture de dépendance avec les rivières *Natashquan* (NAT) et *Romaine* (ROM). Cependant, ces deux dernières rivières ne voient pas leur structure de dépendance changer.

Table 6.6: Statistiques de test et valeurs critiques estimées (triangle supérieur) et années de rupture de la dépendance, s'il y a lieu (triangle inférieur) pour les bassins versants du centre-est du Québec

Bassins	CHU	MOI	ROM	NAT
CHU		0.3251 (0.5247)	0.7173 (0.5769)	0.7419 (0.6217)
MOI			0.5802 (0.4985)	0.4901 (0.6475)
ROM	2035	2027		0.3261 (0.4485)
NAT	2008			

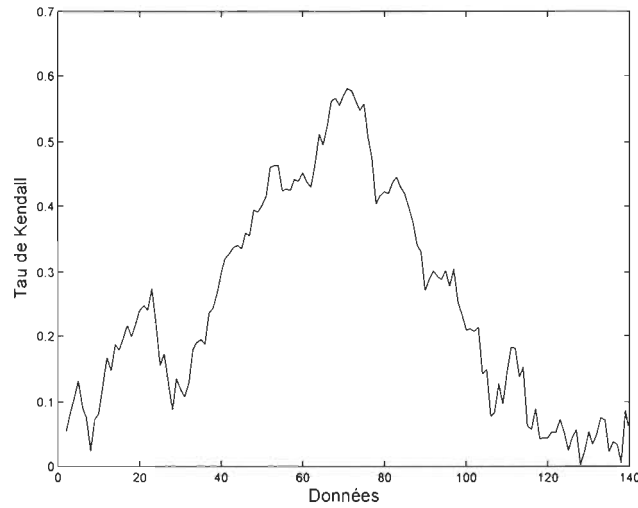


Figure 6.11: Détection graphique de la dépendance pour les rivières *Moisie* (MOI) et *Romaine* (ROM)

6.4 Sommaire des analyses

Ce chapitre a permis de mettre en pratique la nouvelle procédure élaborée au chapitre 4 et dont la puissance a été évaluée au chapitre 5.

La première partie du chapitre a étudié en profondeur le lien de dépendance qui unit les bassins versants de la rivière *À la Baleine* (BAL) et de la rivière *George*. Un changement de moyenne progressif a été observé pour la rivière *À la Baleine* (BAL), alors qu'une modification abrupte de moyenne a été détecté pour la rivière *George*. Cependant, aucune modification de variance n'a été détectée. Par suite de la stabilisation des marges, le test de détection de rupture de dépendance a été appliqué. La statistique de test observée s'est avérée inférieure à la valeur critique estimée, ce qui a permis de conclure que ces deux bassins ont conservé un lien de dépendance identique au cours du temps.

La deuxième partie du chapitre a étudié le lien de dépendance unissant quelques bassins du centre-ouest et du centre-est du Québec. Contrairement aux bassins *À la Baleine* et *George*, plusieurs paires de bassins ont vu leur dépendance se modifier au fil du temps. Ce fut le cas, notamment, pour les rivières *Moisie* et *Romaine*. Dans le futur, il serait intéressant de poursuivre les analyses afin d'explorer les causes possibles de ces changements.

CONCLUSIONS ET PERSPECTIVES

L'étude des points de rupture est un sujet qui a intéressé de nombreux auteurs depuis plusieurs années. Les premiers résultats ont concerné la détection de ruptures abruptes dans la moyenne de séries univariées. Les résultats de Pettitt (1979) sont d'ailleurs une référence quant à l'usage de méthodes non-paramétriques dans ce contexte. Une généralisation de ce modèle a été proposée par Lombard (1987) afin de permettre des ruptures graduelles, ce qui est souvent plus réaliste pour expliquer certains phénomènes climatologiques et hydrologiques. Plusieurs auteurs se sont ensuite intéressés à l'identification de ruptures dans les lois multivariées, par exemple Csörgö & Horváth (1988) et Horváth et al. (1999). Dans tous ces travaux, une approche classique de modélisation simultanée des variables aléatoires est employée. Cette façon de faire est plutôt contraignante puisque les marges sont entièrement dictées par le modèle choisi.

Dans ce mémoire, une approche plus flexible qui consiste à utiliser les copules pour décomposer les comportements marginaux et la dépendance a été adoptée. La méthodologie développée comble un vide dans la littérature statistique car le seul test disponible provient de Dias & Embrechts (2004). Ce test émet cependant beaucoup d'hypothèses peu réalistes pour les applications. L'essentiel de ce mémoire a ainsi consisté à mettre en oeuvre

une nouvelle méthode de détection de points de rupture dans la dépendance, tout en permettant aux marges de se modifier. La statistique de test est non-paramétrique au sens où aucun modèle de copules n'est spécifié et que peu d'hypothèses sur les marges sont émises. Une étude par simulations a montré que la procédure de test est puissante sous de nombreux scénarios. Un outil graphique a également été développé pour identifier la présence d'une rupture dans la dépendance.

Dans le futur, les travaux de ce mémoire pourraient servir de base pour élaborer d'autres méthodes de détection de rupture de dépendance. Une extension naturelle serait basée sur des comparaisons de copules empiriques plutôt que sur des différences de tau de Kendall. Les résultats asymptotiques nécessaires à la validation d'une telle approche nécessitent toutefois une maîtrise de la théorie des processus empiriques. Une autre avenue envisageable concerne le développement d'un modèle à rupture de dépendance graduelle, à l'instar du modèle de Lombard (1987) pour les séries univariées. Celui-ci généraliserait le modèle d'un changement abrupt de dépendance, tel que considéré dans ce travail.

ANNEXE A

PROGRAMMES EN MATLAB

On retrouve dans l'ordre, les programmes utilisés pour effectuer la procédure complète de détection de rupture au niveau de la dépendance. En premier lieu, le programme `testLombard_kd` permet de détecter des ruptures de moyenne et de variance dans les marginales à l'aide du test basé sur l'article de Lombard (1987). Suit le programme `score` qui permet de choisir entre trois différentes fonctions scores pour vérifier la présence de rupture dans la moyenne ou la variance.

Ensuite, les programmes `transformation_mean` et `transformation_var` permettent, s'il y a lieu, de stabiliser, respectivement, la moyenne et la variance des séries.

Finalement, le programme `rupture1` termine la procédure en vérifiant la présence ou non de rupture dans la dépendance. Ce dernier programme est basé sur l'article de Gombay et Horvath (1999).

```

1 function [Tn, an1, an2] = testLombard_kd(X,i,annees)
2
3 % en entrée :
4 % X: vecteur dont on veut calculer les rangs
5 % i: est l'indice qui prend les valeurs de scores - 1 : rupture en moyenne;
6 % 2 : rupture en variance (score de Mood) 3 : rupture en variance (score logarithmique)
7 % annees: le vecteur d'années correspondant au vecteur étudié
8 %
9 % en sortie :
10 % Tn: la statistique de lombard
11 % an1: la 1ere année - début de la période de rupture
12 % an2: la 2e année - fin de la période de rupture
13
14 % Auteurs: M.Champagne & K. Dion
15
16
17 [tri, indice]=sort(X);
18 [tri, Rang]=sort(indice); % rangs du vecteur
19
20 phi = score(i,length(X),Rang); % fonction score i=1: moyenne, i=2: variance (Mood),
    i=3: variance
21
22 phim = mean(phi);
23
24 phiet = std(phi);
25
26 S = (phi - phim)/phiet;
27
28 ml = cumsum(S); %somme les S(i)
29
30 L = zeros(length(X),length(X)); %crée la matrice L(T1,T2)
31 %rempli la matrice L
32 for T1 = 1:length(X)-1
33     for T2 = T1+1:length(X)
34         L(T1,T2) = sum(ml(T1 + 1:T2));
35     end
36 end
37
38 L2 = L.^2; %met chaque élément de la matrice au carré
39
40 Vn= 0;
41 for T1 = 1:(length(X)-1)
42     for T2 = T1+1:length(X)
43         Vn = Vn + L2(T1,T2);
44     end
45 end
46
47 %Division pour utiliser la valeur critique de l'article de Lombard
48 Tn = Vn/(length(X))^5;
49
50 % KD: déterminer t1 et t2
51 mat= zeros(length(X),length(X)); %KD crée la matrice pour déterminer t1 et t2

```

```

52
53     for x = 1:length(X)-1
54         for y = x+1:length(X)
55             u=x/length(X);
56             v=y/length(X);
57             sig2 = (((1-u)^3 * (1 + 3*u))/12) - ((1-v)^3 * (1 + 3*v))/12 ...
58                   - (((1-v)^2) * (v^2 - u^2))/2;
59             mat(x,y) = abs(L(x,y))/sqrt(sig2);
60         end
61         clear u v sig2
62     end
63
64     if Tn > 0.0403
65         [maximum, id] = max(mat);
66         [Lmaximum, t2] = max(maximum);
67         t1=id(t2);
68         clear maxi mat id x y
69     else
70         t1=0; t2=0;
71     end
72
73     if t1==0
74         an1 = NaN;
75         an2 = NaN;
76     else
77         an1 = annees(t1);
78         an2 = annees(t2);
79     end

```

```
1 function phi = score(i,n,R)
2
3 %en entrée :
4 % i: fonction score utilisée
5 % n: le nombre de données de la série
6 % R : rang du vecteur obtenu dans le programme testLombard_kd
7 %
8 %
9 % en sortie:
10 % phi : fonction phi qui permettra de calculer la matrice L dans le programme testLombard_kd
11 %
12
13 % Auteure: M.Champagne
14
15 if i == 1 %fonction score pour un changement dans la moyenne
16
17     u = 1:n;
18
19     phi(u) = 2*(R(u)/(n+1)) - 1;
20
21 else
22     if i == 2 %fonction score de Mood pour un changement dans la variance dans
une distribution symétrique
23
24         u = 1:n;
25
26         phi(u) = (2*(R(u)/(n+1)) - 1).^2;
27
28     else %fonction score logarithmique pour un changement dans la variance sur l'intervalle
(0,inf)
29         u = 1:n;
30
31         phi(u) = log(1-(R(u)/(n+1)));
32     end
33 end
```

```
1 function Y = transformation_mean(X,T1,T2,n)
2
3 %Auteure: M.Champagne
4
5 %Transformation pour stabiliser la moyenne de la série X
6
7 %Trouve les valeurs des moyennes avant T1 et après T2
8
9 epsilon1 = mean(X(1:T1));
10 epsilon2 = mean(X(T2:n));
11
12 %Ramène les marges égales
13
14 Z1 = X(1:T1);
15
16 for k = (T1+1):T2
17     Z1(k) = X(k) - ((k-T1)*(epsilon2-epsilon1)/(T2-T1));
18 end
19
20 Z2 = X(T2+1:n)-(epsilon2 - epsilon1);
21
22 Y = [Z1;Z2];
```

```
1 function Y = transformation_var(X,T1,T2,n)
2
3 %Auteure: M.Champagne
4
5 %Transformation pour stabiliser la variance de la série X
6
7 %Trouvons les valeurs des variances avant T1 et après T2
8
9
10 epsilon1 = std(X(1:T1));
11 epsilon2 = std(X(T2 + 1:n));
12
13 %Ramène maintenant les marges égales
14
15 Z1 = X(1:T1);
16
17 for k = (T1+1):T2
18     Z1(k) = ((T2-T1)*epsilon1*X(k) + (k-T1)*(epsilon2-epsilon1)*mean(X))/((T2-T1)*epsilon1
19     + (k-T1)*(epsilon2-epsilon1));
20 end
21
22 Z2 = ((X(T2+1:n))*epsilon1/epsilon2) + (mean(X)*(epsilon2-epsilon1)/epsilon2);
23
24 Y = [Z1;Z2];
```



```

1  function [Mn,critique] = rupture1(X,N)
2
3  %En entrée:
4  %La matrice X dont on veut vérifier s'il y a un point de rupture et un
5  %nombre N qui permettra de calculer la valeur critique associée à la matrice
6  %X
7  %
8  %En sortie:
9  %Mn, la statistique du test de rupture
10 %critique, la valeur critique associée à la matrice X, selon le test
11 %développé par Gombay et Horváth (1999)
12
13 %Auteur: M.Champagne
14
15
16 n=length(X);
17
18 %Calcul les différences de taus de Kendall et sélectionne la plus élevée
19 vec = zeros(n-3,1);
20 for k=2:n-2
21     vec(k-1) = abs(tau(X(1:k,:),:)-tau(X(k+1:n,:),:))*(k*(n-k)/n^(3/2));
22 end
23 Mn = max(vec); %statistique du test
24
25
26 %Méthode de Gombay et Horváth (1999) pour la valeur critique
27
28 %Matrice d'indicateurs
29 A = zeros(n,n);
30 for i=1:n-1
31     for j=i+1:n
32         if ( (X(i,1)-X(j,1))*(X(i,2)-X(j,2)) > 0 )
33             A(i,j)=1;
34         end
35         A(j,i)=A(i,j);
36     end
37 end
38
39
40 %Calcul des statistiques bootstrap Mn(1), ..., Mn(N)
41 M = zeros(N,1);
42 for p=1:N
43
44     %vecteur de n variables uniformes sur (-sqrt(3),sqrt(3))
45     ep = unifrnd(-sqrt(3),sqrt(3),n,1);
46
47     %calcul de la moyenne de ep de 1 à k
48     m1 = zeros(n);
49     m1=cumsum(ep)./[1:n]';
50
51     %calcul de la moyenne de ep de k+1 à n
52     m2 = zeros(n-1);

```

```

53     for i=2:n
54         m2(i-1) = mean(ep(i:n));
55     end
56
57     %calcul de U(1), ..., U(n-1)
58     U = zeros(n-1,1);
59     for k=1:n-1
60
61         %Calcul de la sommation I
62         part1 = ep(1:k)'*A(1:k,1:k)*ones(k,1);
63         V1 = (n-k)*(part1 - m1(k)*sum(sum(A(1:k,1:k))))/(k*n);
64
65         %Calcul de la sommation II
66         part2 = ep(k+1:n)'*A(k+1:n,k+1:n)*ones(n-k,1);
67         V2 = k*(part2 - m2(k)*sum(sum(A(k+1:n,k+1:n))))/(n*(n-k));
68
69         U(k) = abs(4*(V1-V2)/sqrt(n));
70     end
71
72     M(p) = max(U);
73 end
74
75 %place les Mn en ordre croissant pour sélectionner la 95 ème valeur
76 %ordonnée qui deviendra la valeur critique
77 B = sort(M);
78
79 critique = B(int16(N*0.95));
80
81
82 % Graphique permettant de visualiser les différences de taus de Kendall obtenues
83 % ainsi que la statistique Mn et la valeur critique
84
85 l = 2:n-2;
86 h = 1:n-3;
87 plot(l,vec(h));
88 title('éDtection du point de rupture ');
89 xlabel('Data','FontSize',14);
90 ylabel('Kendall tau ','FontSize',14);
91 hold on; plot(l,critique,'—','color','r'); hold off;

```

Références

- ABD-RABOU, A. E. S. & GAD, A. M. (2007). A new change-point rank tests. *Journal of Data Science* **5**, 379–392.
- ANTOCH, J. & HUŠKOVÁ, M. (1996). Tests and estimators for epidemic alternatives. *Tatra Mt. Math. Publ* **7**, 311–329.
- ANTOCH, J. & HUŠKOVÁ, M. (2001). Permutation tests in change point analysis. *Statistics and Probability Letters* **53**, 37–46.
- BARBE, P., GENEST, C., GHOUDI, K. & RÉMILLARD, B. (1996). On Kendall's process. *Journal of multivariate analysis* **58**, 197–229.
- BERG, D. (2009). Copula goodness-of-fit testing: An overview and power comparison. *The European Journal of Finance* , in press.
- BERKES, I., GOMBAY, E. & HORVÁTH, L. (2009). Testing for changes in the covariance structure of linear processes. *Journal of Statistical Planning and Inference* **139**, 2044–2063.
- BORKOWF, C. B. (2002). Computing the nonnull asymptotic variance and

- the asymptotic relative efficiency of spearman's rank correlation. *Computational Statistics and Data Analysis* **39**, 271–286.
- BRYDEN, E., CARLSON, J. B. & CRAIG, B. (1995). *Some Monte Carlo results on nonparametric changepoint tests*. Federal Reserve Bank of Cleveland, Research Dept.
- CAPÉRAÀ, P., FOUGÈRES, A.-L. & GENEST, C. (2000). Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis* **72**, 30–49.
- CARLSTEIN, E. (1988). Nonparametric change-point estimation. *The Annals of Statistics* **16**, 188–197.
- CAYA, D. & LAPRISE, R. (1999). A semi-implicit semi-Lagrangian regional climate model: The Canadian RCM. *Monthly Weather Review* **127**, 341–362.
- CHEN, J. & GUPTA, A. K. (1999). Change point analysis of gaussian model. *Statistical Papers* **40**, 323–333.
- CSÖRGÖ, M. & HORVÁTH, L. (1988). Invariance principles for changepoint problems. *Journal of Multivariate Analysis* **27**, 151–168.
- DIAS, A. & EMBRECHTS, P. (2004). Change-point analysis for dependence structures in finance and insurance. *Risk Measures for the 21st Century*, 321–335.
- FERGER, D. (1994). On the power of nonparametric changepoint-tests. *Metrika* **41**, 277–292.

- FERGER, D. (1995). Nonparametric tests for nonstandard change-point problems. *The Annals of Statistics* **23**, 1848–1861.
- FERMANIAN, J.-D. (2005). Goodness-of-fit tests for copulas. *Journal of multivariate analysis* **95**, 119–152.
- FISHER, N. I. & SWITZER, P. (1985). Chi-plots for assessing dependence. *Biometrika* **72**, 253–265.
- FISHER, N. I. & SWITZER, P. (2001). Graphical assessment of dependence: Is a picture worth 100 tests ? *The American Statistician* **55**, 233–239.
- FLATO, G. M. & BOER, G. J. (2001). Warming asymmetry in climate change simulations. *Geophysical Research Letters* **28**, 195–198.
- FOTOPOULOS, S. B. & JANDHYALA, V. K. (2007). On Hinkley’s estimator: Inference about the change point. *Statistics & Probability Letters* **77**, 1449–1458.
- GENEST, C. & BOIES, J.-C. (2003). Detecting dependence with Kendall plots. *The American Statistician* **57**, 275–284.
- GENEST, C., GHOUDI, K. & RIVEST, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–552.
- GENEST, C. & MACKAY, R. J. (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *La Revue Canadienne de Statistique* **14**, 145–159.

- GENEST, C., QUESSY, J.-F. & RÉMILLARD, B. (2002). Tests of serial independence based on Kendall's process. *La Revue Canadienne de Statistique* **30**, 441–461.
- GENEST, C., QUESSY, J.-F. & RÉMILLARD, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics* **33**, 337–366.
- GENEST, C. & RÉMILLARD, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. In *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques*, vol. 44.
- GENEST, C., RÉMILLARD, B. & BEAUDOIN, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance : Mathematics and Economics* **44**, 199–213.
- GENEST, C. & RIVEST, L. P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association* **88**, 1034–1043.
- GOMBAY, E. (2008). Change detection in autoregressive time series. *Journal of Multivariate Analysis* **99**, 451–464.
- GOMBAY, E. & HORVÁTH, L. (1999). Change-points and bootstrap. *Environmetrics* **10**.
- HACHÉ, M., PERREAULT, L., RÉMILLARD, L. & BOBÉE, B. (1999). Une approche pour la selection des distributions statistiques: application au bassin hydrographique du Saguenay-Lac St-Jean. *Canadian Journal of Civil Engineering* **26**, 216–225.

- HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1–17.
- HINKLEY, D. V. (1972). Time-ordered classification. *Biometrika* **59**, 509–523.
- HORVÁTH, L., KOKOSZKA, P. & STEINEBACH, J. (1999). Testing for changes in multivariate dependent observations with an application to temperature changes. *Journal of Multivariate Analysis* **68**, 96–119.
- HORVÁTH, L. & STEINEBACH, J. (2000). Testing for changes in the mean or variance of a stochastic process under weak invariance. *Journal of Statistical Planning and Inference* **91**, 365–376.
- HSU, D. A. (1977). Tests for variance shift at an unknown time point. *Applied Statistics* **26**, 279–284.
- HUŠKOVÁ, M. & MEINTANIS, S. G. (2006). Change point analysis based on empirical characteristic functions. *Metrika* **63**, 145–168.
- JANDHYALA, V. K., FOTOPOULOS, S. B. & EVAGGELOPOULOS, N. E. (2000). A comparison of unconditional and conditional solutions to the maximum likelihood estimation of a change-point. *Computational Statistics & Data Analysis* **34**, 315–334.
- JANDHYALA, V. K., FOTOPOULOS, S. B. & HAWKINS, D. M. (2002). Detection and estimation of abrupt changes in the variability of a process. *Computational Statistics & Data Analysis* **40**, 1–19.

- JARUŠKOVÁ, D. (1997). Some problems with application of change-point detection methods to environmental data. *Environmetrics* **8**, 469–483.
- LEE, A. J. (1990). *U-statistics: Theory and Practice*. M. Dekker.
- LILLIEFORS, H. W. (1967). On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association* **62**, 399–402.
- LOMBARD, F. (1987). Rank tests for changepoint problems. *Biometrika* **74**, 615–624.
- MESFIOUI, M., QUESSY, J.-F. & TOUPIN, M.-H. (2009). On a new goodness-of-fit process for families of copulas. *Canadian Journal of Statistics* **37**, 80–101.
- MUSIC, B. & CAYA, D. (2007). Evaluation of the hydrological cycle over the mississippi river basin as simulated by the canadian regional climate model (CRCM). *Journal of Hydrology* **8**, 969–988.
- NELSEN, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer.
- PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–114.
- PERREAULT, L., BERNIER, J., BOBÉE, B. & PARENT, É. (2000a). Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *Journal of Hydrology* **235**, 221–241.

- PERREAULT, L., BERNIER, J., BOBÉE, B. & PARENT, É. (2000b). Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. *Journal of Hydrology* **235**, 242–263.
- PETTITT, A. N. (1979). A non-parametric approach to the change point problem. *Applied Statistics* **28**, 126–135.
- SCARIANO, S. M. & WATKINS, T. A. (1988). Nonparametric point estimators for the change-point problem. *Communications in Statistics-Theory and Methods* **17**, 3645–3675.
- SCINOCCA, J. F., MCFARLANE, N. A., LAZARE, M., LI, J. & PLUMMER, D. (2008). The CCCma third generation AGCM and its extension into the middle atmosphere. *Atmospheric Chemistry and Physics Discussions* **8**, 7883–7930.
- SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.
- VAN DER VAART, A. W. (1998). Asymptotic statistics. Cambridge series in statistical and probabilistic mathematics.
- WANG, W. & WELLS, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association* **95**, 62–76.
- ZOU, C., LIU, Y., QIN, P. & WANG, Z. (2007). Empirical likelihood ratio test for the change-point problem. *Statistics and Probability Letters* **77**, 374–382.